

Faculdade de Engenharia da Universidade do Porto



Estimação de Investimentos na Rede de Distribuição

Pedro Miguel Ferreira Macedo

VERSÃO FINAL

Dissertação realizada no âmbito do
Mestrado Integrado em Engenharia Electrotécnica e de Computadores
Major Energia

Orientador: Prof. Dr. José Nuno Moura Marques Fidalgo

Julho de 2014

A Dissertação intitulada

“Estimação de Investimentos na Rede de Distribuição”

foi aprovada em provas realizadas em 16-07-2014

o júri



Presidente **Professora Doutora Maria Helena Osório Pestana de Vasconcelos**
Professora Auxiliar do Departamento de Engenharia Eletrotécnica e de
Computadores da Faculdade de Engenharia da Universidade do Porto



Professor Doutor Filipe Miguel Tavares Azevedo
Professor Adjunto do Departamento de Engenharia Electrotécnica do Instituto
Superior de Engenharia do Porto



Professor Doutor José Nuno Moura Marques Fidalgo
Professor Associado do Departamento de Engenharia Eletrotécnica e de
Computadores da Faculdade de Engenharia da Universidade do Porto

O autor declara que a presente dissertação (ou relatório de projeto) é da sua exclusiva autoria e foi escrita sem qualquer apoio externo não explicitamente autorizado. Os resultados, ideias, parágrafos, ou outros extratos tomados de ou inspirados em trabalhos de outros autores, e demais referências bibliográficas usadas, são corretamente citados.



Autor - Pedro Miguel Ferreira Macedo

Faculdade de Engenharia da Universidade do Porto

Resumo

Os Operadores da Rede de Distribuição (ORD), de acordo com a regulamentação vigente relativa ao sector elétrico (Decreto Lei n.º 215-A/2012), deparam-se com a obrigatoriedade de elaborar o Plano de Desenvolvimento e Investimento da Rede de Distribuição. Este deve ser projetado para um horizonte temporal de cinco anos e atualizado a cada biénio, tendo por base a caracterização técnica da rede e a sua oferta e procura, atuais e previstas.

Neste plano, o ORD deve incluir os investimentos futuros a realizar na Rede Nacional de Distribuição (RND), os quais se dividem habitualmente em dois grupos, os Investimentos Obrigatórios (IO) e os de iniciativa do ORD. Surge então a necessidade de desenvolver uma metodologia e uma ferramenta que permita ao ORD, de forma prática, estimar os IO e as participações financeiras e em espécie, sendo este o foco do estudo desenvolvido.

A estimação dos IO é feita em função do seu histórico, expectativas de crescimento da carga, índices socioeconómicos, como a taxa de inflação, taxa de desemprego, índice de construção, entre outros. Os IO abrangem sobretudo os investimentos com base nas solicitações de clientes, tanto a nível de novas ligações como de aumento de energia consumida, associadas a operações de expansão ou reforço na RND.

Devido ao escasso histórico existente, a realização das estimações é condicionada por técnicas pouco exigentes em termos de quantidade de dados. São ainda aplicadas diversas transformações aos dados, como normalização, variação anual e dessensibilização.

Por fim, é desenvolvido um programa em Matlab que permite automatizar o processo da realização das estimações, com recurso a uma regressão linear e ao critério dos mínimos quadrados. Este revela-se de extrema importância para a agilização de todo o processo, na medida em que é capaz de avaliar todos os modelos de estimação capazes de serem gerados com as variáveis existentes, para as diversas transformações dos dados.

Abstract

According to the current legislation concerning the electricity sector (Decree Law 215-A/2012), the Distribution System Operators (DSO) are required to prepare a Network Distribution Development and Investment Plan (NDDIP). This should be designed for a time frame of five years and updated every two years, based on the technical characteristics of the network and its current and predicted supply and demand.

In this plan, the DSO must include future investments to be made in the National Distribution Network (NDN), which is usually divided into two groups, the Required Investments (RI) and the investments undertaken by initiative of the operator. Thus, arises the need to develop a methodology and a tool that allows the DSO, in a practical way, to estimate the RI as well as the financial and in-kind contributions associated with them, which is the focus of the study developed.

The estimation of RI is made on the basis of its historical past, expectations of load growth, socio-economic indicators such as the inflation rate, unemployment rate, construction index, among others. The RI mostly covers the investments based on requests from customers, both in terms of new connections as the increment of energy consumption, which can be described by the upgrades carried out in the NDN.

The realization of the estimations is constrained by the lack of existing history. Thus, techniques applied in conventional estimations are tested and evaluated. Various approaches are also applied to data, such as normalization, annual variation and desensitization.

Finally, a Matlab program to automate the process of the realization of estimations, using a linear regression and the least squares criterion is developed. The program is extremely important for the improvement of the whole process, as it is capable of evaluating all estimation models that can be generated with the existing variables for the various approaches to data.

Agradecimentos

Um agradecimento muito especial a todos os meus colegas e amigos que me ajudaram a ultrapassar certas etapas com um excelente clima de amizade, e que de uma ou de outra forma, fizeram com que me lembre de todo este caminho com alegria e positivismo.

Ao professor Doutor José Nuno Fidalgo por toda a orientação, paciência, tranquilidade e tempo disponibilizado.

À Ana Sofia, por ser tão incrível, me apoiar e guiar com tanto carinho e dedicação.

À Irene, pelo brilho que dá a cada novo dia, pela sua paciência e pela forma como encara e me faz encarar a vida.

Aos meus familiares por serem o meu suporte de vida e tornarem possível o meu percurso até este ponto, com um especial obrigado à minha mãe por toda a confiança e apoio demonstrados.

O meu especial obrigado,

Pedro Macedo

“Never underestimate the power of a simple tool”

Craig Bruce

Índice

Resumo	i
Abstract.....	iii
Agradecimentos	v
Índice.....	ix
Lista de figuras	xi
Lista de tabelas	xv
Abreviaturas e Símbolos	xvii
Capítulo 1	1
Introdução.....	1
1.1 - Enquadramento	1
1.2 - Objetivos	3
1.3 - Estrutura do Documento.....	4
Capítulo 2	5
Metodologia	5
2.1 - Caracterização Geral da Informação	6
2.1.1 - Origem dos Dados	7
2.1.2 - Alterações Especiais	18
2.1.3 - Impacto de Alterações na Legislação	21
2.2 - Pré-processamento de Dados	26
2.3 - Processos de Estimação.....	31
2.3.1 - Relação dos Indicadores de Construção e das Realizações com as restantes variáveis	33
2.3.2 - Indicadores de Construção - IC, NNF e NNC	35
2.3.3 - Realizações - NPPT, CNLA e CNCS.....	42
2.3.4 - Investimentos Obrigatórios e Comparticipações	44
2.4 - Programa para automatização dos Processos de estimação	47
Capítulo 3	55
Resultados.....	55
3.1 - Indicadores de Construção, IC, NNF e NNC.....	55
3.2 - Realizações EDP Distribuição: NPPT, CNLA e CNCS	60
3.3 - Investimentos Obrigatórios e Comparticipações	69

Capítulo 4	75
Conclusões e Desenvolvimentos Futuros.....	75
4.1 - Conclusões Finais	75
4.2 - Desenvolvimentos Futuros	76
Referências	79

Lista de figuras

Figura 2.1 - Fluxograma de tópicos expostos na metodologia	6
Figura 2.2 - Histórico referente aos IO, CF e CE	8
Figura 2.3 - Histórico dos IO distribuídos pelas diversas UO	8
Figura 2.4 - Histórico das CF distribuídas pelas diversas UO	9
Figura 2.5 - Histórico das CE distribuídas pelas diversas UO	9
Figura 2.6 - Histórico e previsão dos valores de energia veiculada na rede	10
Figura 2.7 - Histórico e previsão do NTC veiculados na rede	11
Figura 2.8 - Histórico e previsão do PIB.....	13
Figura 2.9 - Histórico e previsão da TInf.....	13
Figura 2.10 - Histórico e previsão da TDes	14
Figura 2.11 - Histórico e previsão do IC.....	15
Figura 2.12 - Histórico e previsão do NNF	16
Figura 2.13 - Histórico e previsão do NNC	16
Figura 2.14 - Histórico e previsão da NPPT.....	17
Figura 2.15 - Histórico e previsão da CNLA.....	17
Figura 2.16 - Histórico e previsão da CNCS.....	18
Figura 2.17 - CE antes e após correção	19
Figura 2.18 - Impacto dos ajustes nos montantes globais dos IO, CF e CE	20
Figura 2.19 - Impacto dos ajustes nos montantes da DRC Sul dos IO, CF e CE	20
Figura 2.20 - Ilustração dos conceitos de ponto de ligação físico e ponto de ligação para efeitos comerciais (Fonte: ERSE: “Ligações às Redes de Energia Eléctrica”, Setembro de 2011, disponível em www.erse.pt).....	22
Figura 2.21 - Evolução do CAPEX das CF e dos IO decorrente da alteração na legislação, no período referente ao histórico.....	23
Figura 2.22 - Exemplo da composição de uma curva de adaptação, através da composição de sigmóides que modelizam o impacto de alterações na legislação	26

Figura 2.23 - Ilustração da aplicação do método da variação anual dos dados na variável TDes	29
Figura 2.24 - Ilustração da aplicação da dessensibilização nos dados da variável PIB	30
Figura 2.25 - Estimções para a CNLA obtidas pelos ME(EnMTBT, PIB, NNF) para dados de Base, Variação e Dessensibilizados	31
Figura 2.26 - Ilustração de um excerto de uma série com explicação dos valores afetos ao cálculo de R^2 (Fonte: ERSE: “Ligações às Redes de Energia Eléctrica”, Setembro de 2011, disponível em www.erse.pt)	34
Figura 2.27 - Relação, dada por R^2 , entre as variáveis afetas ao setor da construção e às realizações, com as restantes variáveis disponíveis	34
Figura 2.28 - Análise da inclusão de valores históricos de uma variável na sua previsão.....	36
Figura 2.29 - Estimções resultantes da aplicação do ME (NTC, PIB, TDes) às diversas regressões	37
Figura 2.30 - Avaliação das estimções resultantes da aplicação do ME (NTC, PIB, TDes) às diversas regressões	38
Figura 2.31 - Ilustração exemplo da aplicação do método dos mínimos quadrados na estimação do NNF.	39
Figura 2.32 - Ilustração exemplo da aplicação do método dos mínimos quadrados na estimação do NNC.....	39
Figura 2.33 - Relação linear entre o NNF e o IC.....	40
Figura 2.34 - Relação linear entre o NNC e o IC.....	40
Figura 2.35 - Relação exponencial entre o NNF e o IC.	41
Figura 2.36 - Relação exponencial entre o NNC e o IC.....	41
Figura 2.37 - Relação logarítmica entre o NNF e o IC.	41
Figura 2.38 - Relação logarítmica entre o NNC e o IC.....	41
Figura 2.39 - Relação polinomial de 2º grau entre o NNF e o IC.	41
Figura 2.40 - Relação polinomial de 2º grau entre o NNC e o IC.	41
Figura 2.41 - Relação potencial entre o NNF e o IC.	42
Figura 2.42 - Relação potencial entre o NNC e o IC.	42
Figura 2.43 - Estimção das realizações com base na tendência de variação do histórico.....	43
Figura 2.44 - Relação linear, dada por R^2 , entre as variáveis afetas às realizações e os indicadores relativos à construção civil.	43
Figura 2.45 - Relação, dada por R^2 , entre as variáveis disponíveis e os IO.....	45
Figura 2.46 - Sigmoide decrescente de adaptação às alterações na legislação em 2007 para as CF e CE.	46

Figura 2.47 - Curva de adaptação às alterações na legislação para as CF e CE.	47
Figura 2.48 - Fluxograma de processos compreendidos no programa desenvolvido em Matlab.	48
Figura 2.49 - Exemplo dos ME filtrados da MR ordenada, referente à estimação da CNCS, para uma margem com $k=1$ e $K=2$	53
Figura 3.1 - Estimações obtidas pelos dez primeiros modelos, devolvidos pela MR ordenada, referente ao IC.	56
Figura 3.2 - Estimações obtidas pelos dez primeiros ME, devolvidos pela MR ordenada e filtrada para $k=1$, referente ao IC.	56
Figura 3.3 - NRMSE para a média das estimações, fornecidas pelos dez primeiros ME presentes MR ordenada e filtrada, para $k=1$ e $k=2$, na estimação do IC.	57
Figura 3.4 - Estimações com recurso às diversas regressões para o NNF.	58
Figura 3.5 - Estimações com recurso às diversas regressões para o NNC.	59
Figura 3.6 - NRMSE obtida nas estimações do NNF, pelas diversas regressões aceitáveis.	59
Figura 3.7 - NRMSE obtida nas estimações do NNC, pelas diversas regressões aceitáveis.	60
Figura 3.8 - Estimações obtidas pelos dez primeiros modelos, devolvidos pela MR ordenada, referente à NPPT.	61
Figura 3.9 - Estimações obtidas pelos dez primeiros ME, devolvidos pela MR ordenada e filtrada para $k=2$, referente à NPPT.	62
Figura 3.10 - Estimações obtidas pelos dez primeiros ME, devolvidos pela MR ordenada e filtrada para $k=1$, referente à NPPT.	62
Figura 3.11 - NRMSE para a média das estimações, fornecidas pelos dez primeiros ME presentes MR ordenada e filtrada, para $k=1$ e $k=2$, na estimação da NPPT.	63
Figura 3.12 - Estimações obtidas pelos dez primeiros modelos, devolvidos pela MR ordenada, referente à CNLA.	64
Figura 3.13 - Estimações obtidas pelos dez primeiros ME, devolvidos pela MR ordenada e filtrada para $k=2$, referente à CNLA.	65
Figura 3.14 - Estimações obtidas pelos dez primeiros ME, devolvidos pela MR ordenada por uma média ponderada calculada com apenas 20% de teste, referente à CNLA.	65
Figura 3.15 - NRMSE para a média das estimações fornecidas pelos dez melhores ME obtidos pela MR ordenada, para as diferentes formas de cálculo da média ponderada, referente às CNLA.	66
Figura 3.16 - Estimações obtidas pelos dez primeiros ME, devolvidos pela MR ordenada e filtrada para $k=2$, referente à CNCS.	67
Figura 3.17 - Estimações obtidas pelos dez primeiros ME, devolvidos pela MR ordenada e filtrada para $k=1$, referente à CNCS.	67
Figura 3.18 - NRMSE para a média das estimações, fornecidas pelos dez primeiros ME presentes MR ordenada e filtrada, para $k=1$ e $k=2$, na estimação do CNCS.	68

Figura 3.19 - Estimções obtidas pelos dez primeiros modelos, devolvidos pela MR ordenada, referente aos IO.	69
Figura 3.20 - Estimções obtidas pelos dez primeiros ME, devolvidos pela MR ordenada e filtrada para $k=1$, referente aos IO.	70
Figura 3.21 - NRMSE para a média das estimções, fornecidas pelos dez primeiros ME presentes MR ordenada e filtrada, para $k=1$ e $k=2$, na estimção dos IO.	70
Figura 3.22 - Histórico e Previsão dos IO.	71
Figura 3.23 - Histórico e previsão das CF e CE	72
Figura 3.24 - Evolução do CAPEX das CF e dos IO decorrente da alteração na legislação	73

Lista de tabelas

Tabela 2.1 - Fundos Europeus destinados à eletrificação da Serra da Serpa.....	20
Tabela 2.2 - Exemplo da aplicação do método Min-Máx às variáveis PIB e TDes	28
Tabela 2.3 - Exemplo da aplicação da dessensibilização à variável PIB	30
Tabela 2.4 - Análise da inclusão de valores históricos de uma variável na sua previsão	37
Tabela 2.5 - Erros dos modelos usados na estimação, com recurso ao método dos mínimos quadrados, do NNF e NNC	40
Tabela 2.6 - Ilustração exemplo da correspondência de entradas dessensibilizadas com as saídas.....	49
Tabela 2.7 - Exemplo da posição na matriz de entrada, das colunas importadas para o ME...	50
Tabela 2.8 - Legenda do código referente ao tipo de ME criado.	50
Tabela 2.9 - Ilustração exemplo da MR.	51
Tabela 3.1 - ME usados para obtenção da estimação final do IC.	58
Tabela 3.2 - ME usados para obtenção da estimação final da NPPT.	63
Tabela 3.3 - ME usados para obtenção da estimação final da CNLA.	66
Tabela 3.4 - ME usados para obtenção da estimação final da CNCS.	68
Tabela 3.5 - ME usados para obtenção da estimação final dos IO.....	71

Abreviaturas e Símbolos

AT	Alta Tensão
BT	Baixa Tensão
CAPEX	<i>Capital Expenditure</i>
CE	Comparticipações em Espécie
CF	Comparticipações Financeiras
CNCS	Construção de Novos Cabos Subterrâneos
CNLA	Construção de Novas Linhas Aéreas
CS	Cabos Subterrâneos
DAT	Direção de Alta Tensão
DGEG	Direção Geral de Energia e Geologia
DPC	Direção de Projeto e Construção
DRC	Direção de Redes e Clientes
EnMTBT	Energia Utilizada na Rede de MT e BT
EnT	Energia Total Veiculada na Rede
ERSE	Entidade Reguladora do Sistema Energético
FEUP	Faculdade de Engenharia da Universidade do Porto
IC	Índice de Construção
LA	Linhas Aéreas
INE	Instituto Nacional de Estatística
IO	Investimentos Obrigatórios
IHPC	Índice Harmonizado de Preços ao Consumidor
IPC	Índice de Preços ao Consumidor
ME	Modelo de Estimação
MR	Matriz Resultados
MT	Média Tensão
NNC	Nº de Novas Construções
NNF	Nº de Novos Fogos Construídos
NPPT	Nova Potência instalada na rede, em termos de PT
NTC	Nº Total de Clientes Veiculados na Rede
ORD	Operadores da Rede de Distribuição
PDIRD	Plano de Desenvolvimento e Investimento da Rede de Distribuição
PDIRT	Plano de Desenvolvimento e Investimento da Rede de Transporte
PIB	Produto Interno Bruto

PT	Posto de Transformação
RND	Rede Nacional de Distribuição
RNT	Rede Nacional de Transporte
RRC	Regulamento de Relações comerciais
TInf	Taxa de Inflação
TDes	Taxa de Desemprego
UO	Unidade Organizacional
GWh	Unidade de Energia: giga watt-hora
kVA	Unidade de Potencia aparente: quilo volt-ampere
m	Unidade de Comprimento: metro
M€	Unidade Monetária: milhões de euros

Capítulo 1

Introdução

O trabalho de investigação descrito neste documento foi desenvolvido no âmbito da disciplina Dissertação do Mestrado Integrado em Engenharia Electrotécnica e de Computadores da Faculdade de Engenharia da Universidade do Porto.

Devido à obrigatoriedade dos Operadores da Rede de Distribuição (ORD) elaborarem o Plano de Desenvolvimento e Investimento das Redes de Distribuição (PDIRD), onde têm de incluir os investimentos futuros a realizar na Rede Nacional de Distribuição (RND), surge a motivação em desenvolver uma metodologia e uma ferramenta que lhes permita, de forma prática, estimar parte destes investimentos, isto é, os Investimentos Obrigatórios (IO). Os IO abrangem sobretudo os investimento com base nas solicitações de clientes, tanto a nível de novas ligações como de aumento de energia consumida.

Neste capítulo é feita uma abordagem introdutória ao projeto e são definidos os conceitos gerais afetos ao tema desenvolvido. No enquadramento serão aprofundadas as motivações que levaram ao desenvolvimento deste estudo, ao qual se segue uma exposição dos objetivos a realizar. Por fim, será apresentada a estrutura da dissertação.

1.1 - Enquadramento

De acordo com o estipulado no art.º 41º do Decreto-Lei n.º 29/2006, de 15 de Fevereiro [1], que sofreu, mais recentemente, a quinta alteração pelo Decreto-Lei n.º 215-A/2012 [2], os ORD devem elaborar o PDIRD, de forma a assegurar o desenvolvimento adequado e eficiente das redes, no âmbito do mercado interno da eletricidade, tendo em conta a existência de capacidade, por parte das mesmas, para a receção e entrega de eletricidade com níveis adequados de qualidade de serviço e segurança.

O art.º 40º do Decreto-Lei n.º 215-B/2012 [3], que representa a sexta e mais recente alteração do Decreto-Lei n.º 172/2006 [4], referencia que o PDIRD é um plano quinquenal a ser

atualizado de dois em dois anos, o qual deve ser apresentado nos anos pares. No mesmo artigo, é referido ainda que este deve ser compatível com o Plano de Desenvolvimento e Investimento das Redes de Transporte (PDIRT) e incluir a identificação dos principais desenvolvimentos futuros da expansão da rede.

Note-se que, de acordo com o descrito no art.º 41º do Decreto-Lei n.º 215-A/2012, o PDIRD terá de ser aprovado pelo membro do Governo responsável pela área da energia pertencente à Direção Geral de Energia e Geologia (DGEG), após parecer da Entidade Reguladora do Sistema Energético (ERSE) e do operador da Rede Nacional de Transporte (RNT), para, posteriormente, ser submetido a consulta pública.

No PDIRD é descrita a política de investimentos para o período em questão, que podem ser divididos em dois grupos: os IO e os investimentos de iniciativa do ORD. Os IO, foco do trabalho desenvolvido, englobam os investimentos na rede de Baixa Tensão (BT), os quais não se inserem no PDIRD. Estes caracterizam-se pelas obras inerentes à ligação de novos clientes e produtores, incluindo loteamentos e urbanizações, as quais abrangem as obras necessárias à criação de condições na rede para a alimentação dos consumos e as relacionadas com o cumprimento das obrigações estabelecidas nos contratos de concessão [5].

De um modo geral, os IO estão relacionados com os seguintes tipos de ações:

- Ligação de clientes;
- Investimento obrigatório decorrente da ligação de clientes;
- Ligação de urbanizações e loteamentos;
- Remodelação de redes para aumento da vida útil;
- Ligação de PRE.

Os IO são habitualmente previstos segundo estimativas de ligações a efetuar com a RND. Ao analisar a RND de jusante para montante, é de fácil perceção que o aumento de ligações e de carga nas redes BT originam, por vezes, o reforço dos Postos de Transformação (PT), tanto em potência como em número, o que, por sua vez, ocasiona o reforço das redes de Média Tensão (MT). No entanto, na rede de Alta Tensão (AT), incluindo também as subestações, o investimento não ocorre somente pelo crescimento da carga por parte dos clientes AT, MT e BT, mas também pelo aumento das ligações a novas instalações produtoras.

Existem diversos fatores que interferem potencialmente na variação dos montantes referentes aos IO a realizar em cada ano. Estes fatores estão, geralmente, relacionados com o desenvolvimento socioeconómico, como, por exemplo, a redução de atividade no setor industrial ou da construção civil, com alterações de condições locais que impliquem a necessidade de alterar a rede existente, como por exemplo a implantação de uma unidade industrial ou de uma nova urbanização. Junto a estes fatores, existe ainda a alteração da regulamentação, como, por exemplo, a criação de barreiras ou oportunidades para a efetivação de novas ligações, que além de afetar os IO pode afetar também diretamente as Comparticipações Financeiras (CF) e as Comparticipações em Espécie (CE).

Entenda-se as CF como o montante pago pelo requisitante à EDP Distribuição, proveniente dos encargos decorrentes da nova ligação, e as CE como os elementos que após validação da nova ligação passam a constituir parte integrante da RD. Estes conceitos, bem como a forma de cálculo dos montantes, encontram-se detalhados na subsecção 2.1.3.

Devido à importância dada pela EDP Distribuição ao tema, foi estabelecido um protocolo de colaboração entre esta e o INESC Porto, com o objetivo de implementar procedimentos para a estimação dos valores dos IO, incluindo as CF e as CE, a realizar nas redes de distribuição de energia elétrica [6].

Na sequência do projeto de investigação desenvolvido no INESC Porto, surge o tema da dissertação desenvolvido e aqui descrito, proposto pelo orientador Prof. Dr. José Nuno Fidalgo, um dos investigadores do INESC Porto intervenientes no projeto. Na reunião de apresentação do tema com o orientador, dia 12 de Fevereiro de 2014, foram apresentados e disponibilizados dados relativos ao projeto desenvolvido pelo INESC Porto, com o intuito de analisar os procedimentos habitualmente utilizados e de implementar um processo que permita melhorar ou robustecer as estimações solicitadas.

1.2 - Objetivos

O principal objetivo do trabalho consiste em fornecer uma metodologia, com uma ferramenta associada, que permita aos ORD obter as estimações para os IO, CF e CE, para um horizonte temporal que contemple, no mínimo, os 5 anos previstos no PDIRD, com especial enfoque nas estimativas referentes ao primeiro biénio. As estimações podem ser observadas globalmente, o que representa o total referente à EDP Distribuição, ou distribuídas pelas diversas unidades organizacionais (UO) pertencentes a esta, tais como Direção de Projeto e Construção (DPC), Direção de Alta Tensão (DAT) e as seis Direções de Redes e Clientes (DRC) (Norte, Porto, Mondego, Tejo, Lisboa e Sul).

Por forma a responder adequadamente ao principal desafio deste trabalho, é essencial traçar objetivos preliminares, como a caracterização geral da informação disponibilizada. Esta caracterização compreende uma análise das variáveis explicativas, tanto de variáveis relativas à EDP Distribuição como de variáveis socioeconómicas, com o intuito de avaliar a relação destas com as variáveis a prever e a possibilidade de inclusão de novas variáveis.

Depois de analisadas as variáveis explicativas, é necessário, sempre que não tenham sido disponibilizadas pela EDP Distribuição, realizar pesquisas em fontes alternativas ou mesmo fazer previsões das mesmas, o que exige novamente uma análise de relação entre estas e as restantes variáveis já existentes.

1.3 - Estrutura do Documento

Este documento, que visa expor o estudo desenvolvido, com intuito de responder aos desafios anteriormente expostos, encontra-se dividido em quatro capítulos. Abaixo, de forma sucinta encontra-se descrito o que se aborda em cada um destes.

Neste capítulo é feito um enquadramento ao tema, onde são expostos os conceitos gerais associados a este e as motivações que levaram à realização deste estudo.

O capítulo 2 compreende toda a metodologia do trabalho que se caracteriza por ser extenso, o que levou à sua divisão em diversas secções e subsecções. Primeiramente, é feita uma caracterização geral da informação, na qual se descreve sucintamente as variáveis e a origem dos seus dados, que permite ainda observar quais os dados relativos a previsões em falta. No mesmo subcapítulo, são ainda descritas e justificadas quais as alterações excepcionais, necessárias realizar, em alguns dados históricos e avaliados os possíveis impactos de alterações na legislação. Noutra secção, é feito um pré-processamento dos dados, onde se aplicam diversas abordagens, como normalização, variação anual e dessensibilização dos dados. No subcapítulo posterior, são descritos os processos de estimação para as diferentes variáveis que possuem dados de previsão em falta, na qual se analisa ainda a relação entre estas e as restantes variáveis. Posteriormente, é descrito o programa implementado para a automatização dos processos de estimação.

No capítulo 3 são apresentadas as estimações finais para todas as variáveis de estimar, bem como os Modelos de Estimação (ME) que as permitiram obter. São ainda expostas as medidas de erro associadas a cada estimação e retiradas algumas conclusões após uma breve análise.

Por fim, no sexto capítulo serão expostas as conclusões finais e descritos quais os desenvolvimentos futuros.

Capítulo 2

Metodologia

A estimação de IO e das comparticipações é um tema relativamente recente, que surgiu pela necessidade dos ORD terem a obrigatoriedade de as incluírem no PDIRD. Esta necessidade deu origem a estudos de investigação, num projeto desenvolvido no INESC Porto [6] e motivou os desenvolvimentos apresentados nesta dissertação.

Contudo, visto ser um tema tão recente não foi possível encontrar bibliografia sobre o assunto, concluindo-se que não existirão metodologias definidas para aplicar neste tipo de estudos, que se caracteriza por ser extremamente específico. Isto significa que não existem modelos a seguir, pelo que foi necessário testar e experimentar diversos processos, de modo a perceber os que melhor podem constituir uma metodologia eficaz.

Começou-se por experimentar algumas técnicas aplicadas em estimações convencionais mas que frequentemente conduzem a bons resultados. Em seguida foram testadas outras alternativas. As opções tomadas para decidir entre as técnicas a usar na solução final foram guiadas pelo desempenho obtido em cada caso, como se explica nas secções seguintes.

O seguinte fluxograma sintetiza a metodologia, mostrando os principais tópicos abordados neste capítulo e a forma como estes se sucedem.

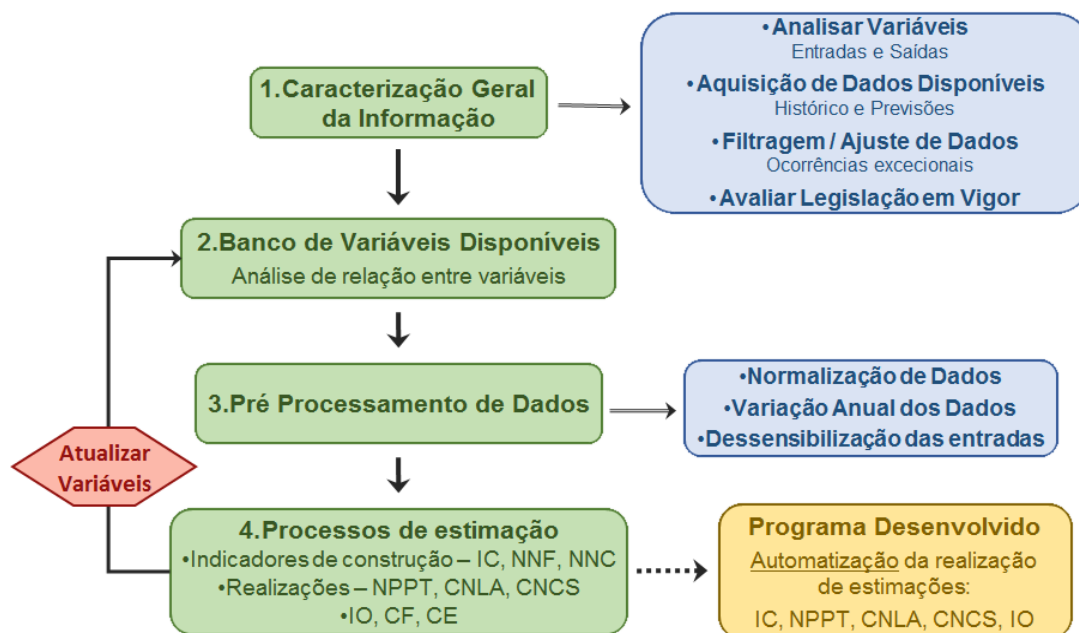


Figura 2.1 - Fluxograma de tópicos expostos na metodologia

2.1 - Caracterização Geral da Informação

Num processo de estimação é necessário previamente analisar toda a informação disponível. Esta informação contempla não só os dados disponibilizados, como todos os detalhes que podem afetar a veracidade dos mesmos ou mesmo fornecer informação sobre como estes se comportam ao longo do histórico.

Além da identificação dos dados disponíveis, constituídos não só pelo histórico mas também por indicadores socioeconómicos, é necessário observar como estes se subdividem. Numa primeira abordagem é feita a divisão entre entradas e saídas disponíveis, que se encontra a seguir descrita:

Dados de Entrada:

- Energia Total Veiculada na Rede (EnT) (GWh);
- Energia Utilizada na Rede de MT e BT (EnMTBT) (GWh);
- Nº Total de Clientes Veiculados na Rede (NTC) (nº);
- Produto Interno Bruto (PIB) (M€);
- Taxa de Inflação (TInf) (%);
- Taxa de Desemprego (TDes) (%);
- Índice de Construção (IC);
- Nº de Novos Fogos Construídos (NNF) (nº);
- Nº de Novas Construções (NNC) (nº);
- Nova Potência instalada na rede, em termos de PT (NPPT) (kVA);
- Construção de Novas Linhas Aéreas (CNLA) (m);
- Construção de Novos Cabos Subterrâneos (CNCS) (m).

Dados de Saída:

- Investimento Obrigatório (IO) (M€);
- Participações Financeiras (CF) (M€);
- Participações em Espécie (CE) (M€).

As variáveis de saída estão subdivididas por oito UO, o qual dá resposta a um dos principais objetivos do tema desenvolvido, apresentando-se distribuídas da seguinte forma:

- EDP Distribuição (global);
- DPC;
- DAT;
- DRC (Norte, Porto, Mondego, Tejo, Lisboa, Sul).

Algumas das variáveis de entrada detêm também a possibilidade de ser observadas/obtidas ao nível de detalhe das DRC. Estas são as que representam os indicadores de construção, fornecidos pelo NNF e NNC, com exceção do IC que só é apresentado para valores globais, e as realizações por parte da EDP Distribuição, como NPPT, CNLA e CNCs.

2.1.1 - Origem dos Dados

Os dados requeridos nos processos de estimação, referentes tanto ao histórico como a previsões, têm diferentes origens, que serão detalhadas nesta secção. Os dados históricos relativos às saídas foram cedidos pela EDP Distribuição e completam um histórico de dados anuais, tendo sido cedidos para os IO os compreendidos entre 2002 e 2012 e para as CF e CE entre 2006 e 2012. Com isto, percebe-se que o histórico é extremamente reduzido, o que constitui uma das principais barreiras na estimação de valores, descartando, desde logo, o recurso a determinadas metodologias exigentes em termos de quantidade de dados, como por exemplo, o recurso a redes neurais [7].

Ao longo de todo o trabalho desenvolvido, é importante referir que as variáveis de entrada são consideradas independentes umas das outras, pelo que não será avaliada a possibilidade de relação e dependência entre estas. No entanto, é essencial que estas não apresentem uma relação direta entre si, ou seja que não sejam redundantes, para que as estimações possam ser realizadas recorrendo ao maior número possível de variáveis.

As saídas serão estimadas até ao ano de 2019, o que perfaz um total de sete anos de previsão, ainda que o PDIRD só exija cinco anos. Isto deve-se, sobretudo, ao facto da atualização de dados históricos para 2013, por parte da EDP distribuição, estar pendente, ficando ainda um ano de margem para possíveis antevisões.

Note-se que devido a uma alteração especial acordada com a EDP Distribuição, que será abordada na subsecção 2.1.2, os montantes referentes às CE aqui apresentados já se encontram corrigidos, não correspondendo ao histórico inicialmente fornecido. De referir ainda que os

valores para a DRC Sul, e consequentemente os globais, alusivos aos IO e CF, entre 2010 e 2013, se encontram subtraídos dos montantes recebidos por fundos europeus destinados à eletrificação da Serra da Serpa, também abordado subsecção 2.1.2.

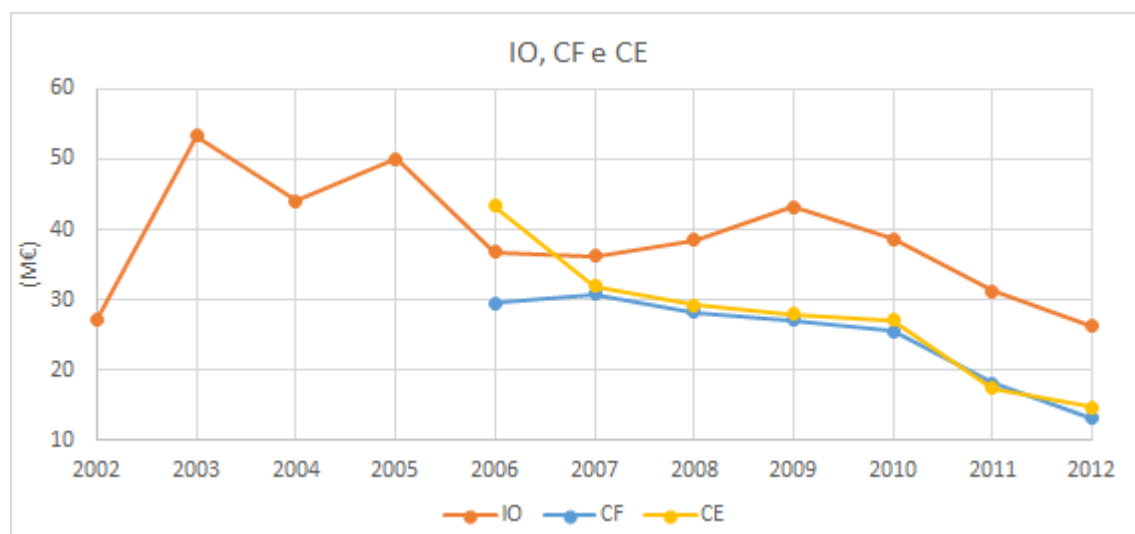


Figura 2.2 - Histórico referente aos IO, CF e CE

Saliente-se que os valores apresentados não correspondem aos valores reais. Por motivos de confidencialidade dos montantes referentes aos IO, CF e CE fornecidos pela EDP, procedeu-se à multiplicação destes por uma constante.

Foram cedidos também pela EDP Distribuição, os valores referentes às variáveis apresentadas para todas as UO, já descritas anteriormente. Os valores correspondentes a esta distribuição para cada uma das variáveis, podem ser observados abaixo. Note-se que os valores referentes à DAT se encontram no eixo da direita para que seja possível a sua visualização, pois estes representam uma parcela mínima em relação às restantes UO.

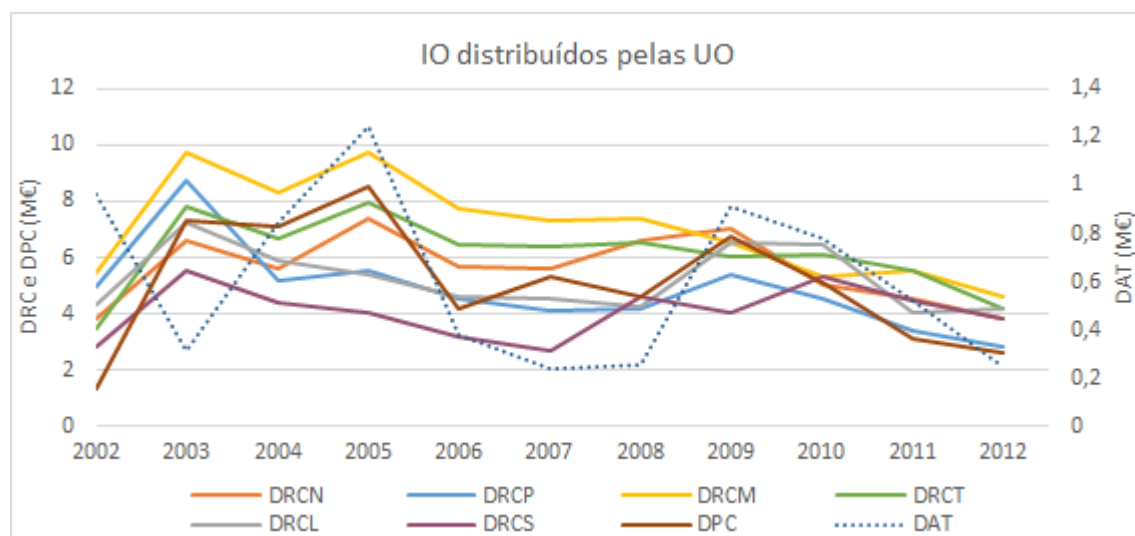


Figura 2.3 - Histórico dos IO distribuídos pelas diversas UO

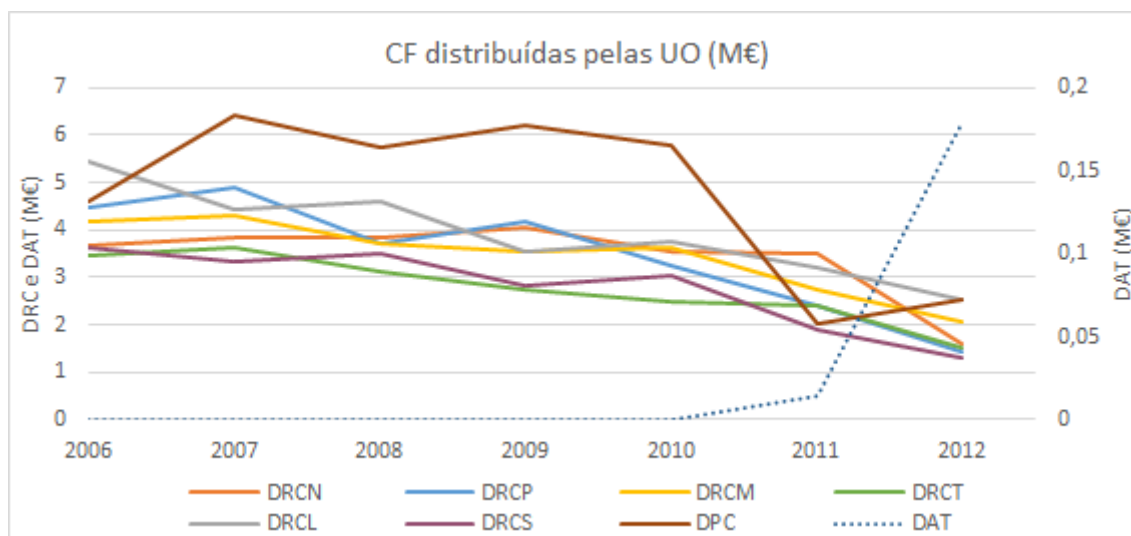


Figura 2.4 - Histórico das CF distribuídas pelas diversas UO

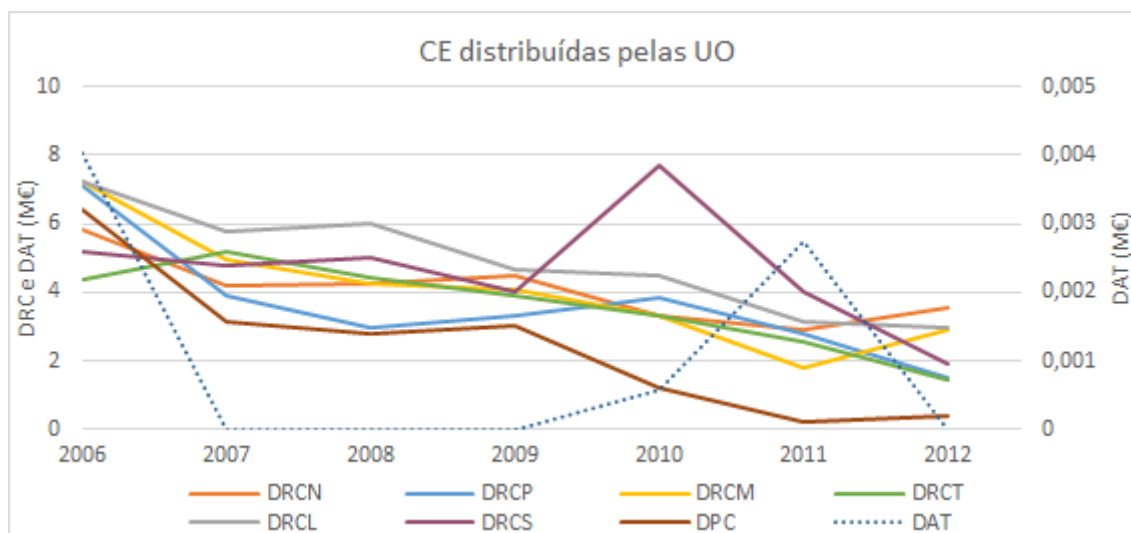


Figura 2.5 - Histórico das CE distribuídas pelas diversas UO

A DAT não só representa uma parcela mínima das UO, como existem anos em que o seu valor é nulo, o que acarreta dificuldades e uma maior imprecisão quanto à sua estimação.

Os dados referentes às restantes variáveis serão preenchidos, sempre que não existam valores históricos, com as previsões já realizadas por instituições credibilizadas. No entanto, sempre que estas previsões ainda não tenham sido realizadas, ou não estejam acessíveis, serão realizadas por processos de estimação descritos posteriormente.

Em seguida, serão apresentados os valores referentes ao histórico e às previsões das variáveis de entrada, utilizadas nos processos de estimação.

Energia

A EDP Distribuição disponibilizou um documento no qual vinham discriminados os valores históricos de 2002 a 2012, assim como as previsões até 2019, da energia veiculada na rede para os diversos níveis de tensão. Desse documento, extraíram-se duas variáveis de entrada: o total e a que agrega a energia em BT e MT, designadas anteriormente como EnT e EnMTBT. Esta divisão surge pelo facto da EnMTBT representar uma parte considerável da energia presente na rede, bem como das realizações afetas a estes níveis de tensão, nomeadamente, de NPPT, CNLA e CNCS. Presume-se que a inclusão desta divisão nos processos de estimação das realizações seja benéfica para a precisão dos resultados.

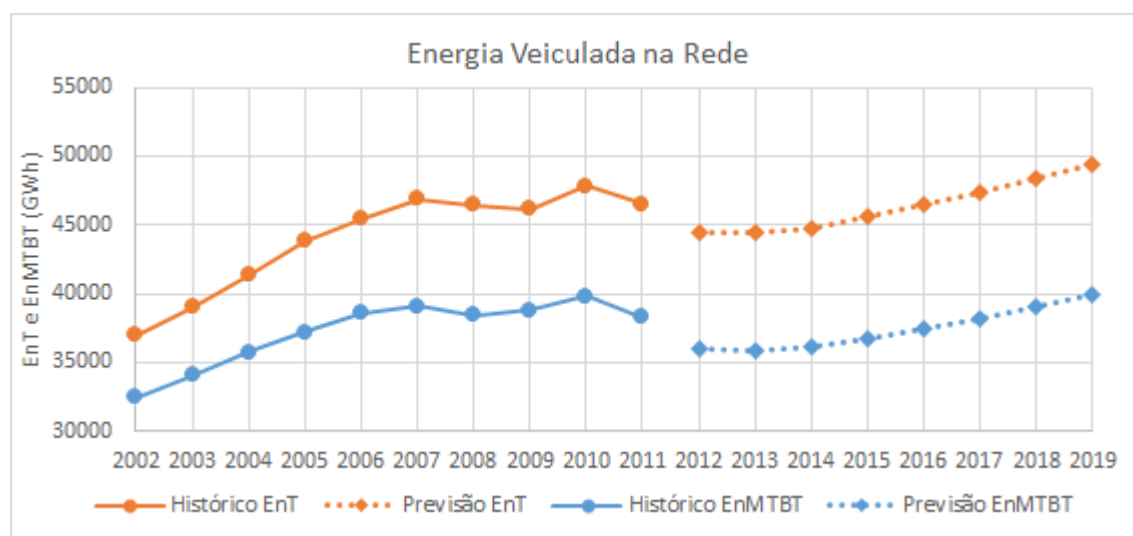


Figura 2.6 - Histórico e previsão dos valores de energia veiculada na rede

A dependência de energia elétrica por parte da população é cada vez maior, pelo que o seu histórico tem seguido um padrão de consumo crescente. No entanto, esta tendência de subida não se fez sentir a partir de 2006, data em que se observa uma estagnação da energia veiculada na rede, seguida de uma queda nos anos mais recentes, possivelmente explicada pela crise económica que se faz sentir no país [8]. Já quanto às previsões, principalmente para os anos posteriores a 2014, observa-se, novamente, uma tendência de subida a um ritmo sensivelmente menor ao observado no histórico anterior a 2006, que poderá ser justificado pela crescente consciencialização da população no que diz respeito ao consumo da energia elétrica, assim como pela utilização de equipamentos cada vez mais eficientes.

Nº de Clientes

Esta variável fornece informação sobre o NTC. Os valores apresentados na figura abaixo, referentes ao histórico e às previsões, foram cedidos pela EDP Distribuição de forma análoga à energia veiculada na rede, anteriormente exposta.

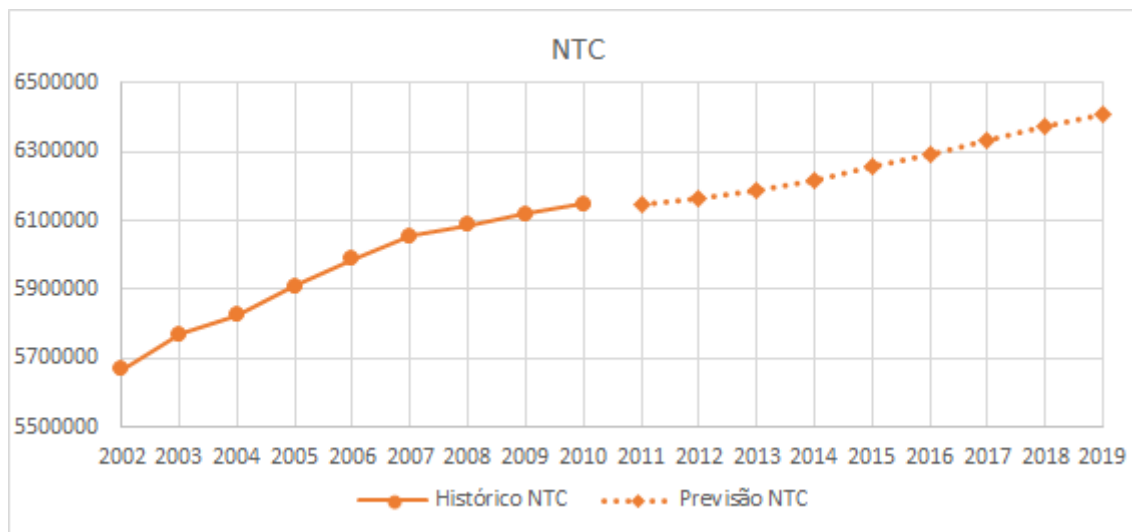


Figura 2.7 - Histórico e previsão do NTC veiculados na rede

Na Figura 2.7 observa-se que o número total de clientes segue uma tendência crescente, com um ligeiro abrandamento entre o período de 2007 a 2013. Este pode ser também explicado pela crise económica que decorre no país e no setor da construção civil, onde se verifica um acentuado decréscimo do NNF, representado na Figura 2.12, sendo, possivelmente, o indicador mais relacionado com o número de novos clientes.

Além disso, em 2007 ocorre uma alteração na legislação relativa às ligações à rede de energia elétrica, exposta na subsecção 2.1.3, que representa um possível impacto no número de novas ligações à rede.

Índices Macroeconómicos - PIB, TInf e TDes

As variáveis relativas a índices macroeconómicos, apresentadas de seguida, foram o PIB, a TInf e a TDes.

Os valores históricos, disponíveis até ao ano de 2013, foram obtidos recorrendo à base de dados da página *online* do Instituto Nacional de Estatística (INE) [9], que disponibiliza os dados num ficheiro Excel com um quadro resumo sobre os dados extraídos, designado pelo INE por “metainformação”. Estes ficheiros podem ser observados em formato eletrónico no CD que acompanha a dissertação, incluídos numa pasta designada “Base para Dados”.

O PIB foi extraído para preços de mercado, também designado por preços correntes, na Base 2006 em M€ para dados anuais, com última atualização a 11 de Fevereiro de 2014.

A TInf pode ser obtida pelo Índice de Preços ao Consumidor (IPC) ou pelo Índice Harmonizado de Preços ao Consumidor (IHPC). Este último foi criado especialmente com base no IPC, de forma a ser possível a comparação de valores entre os restantes países pertencentes

à União Europeia [10]. Saliente-se que a TInf aqui apresentada foi obtida com base na variação do IHPC, devido ao facto das previsões disponíveis para esta variável se fundamentarem neste.

A TInf foi adquirida pela taxa de variação homóloga do IHPC, na Base de 2005 em percentagem, com uma periodicidade mensal, com última atualização a 13 de Maio de 2014. Como a periodicidade é mensal procedeu-se à média aritmética de cada ano, sendo o resultado desta média os valores históricos finais para a TInf.

A TDes foi adquirida com uma periodicidade anual, com base na Série de 1988 e de 2011 em percentagem, para os períodos entre 2002 a 2010 e 2011 a 2013, datando de última atualização a 12 de Julho de 2011 e a 5 de Fevereiro de 2014, respetivamente.

As previsões de todas as variáveis macroeconómicas descritas foram adquiridas com base num documento publicado a 28 de Fevereiro de 2014, pelo Gabinete de Estratégia e Estudos do Ministério da Economia [11]. Este apresenta um quadro resumo que contém as previsões disponibilizadas em estudos por diferentes entidades, nomeadamente, do Fundo Monetário Internacional (FMI), da organização para a Cooperação e Desenvolvimento Económico (OCDE), da Comissão Europeia (CE), do Banco de Portugal (BdP), do Orçamento de Estado para 2014 (OE-2014) e do Documento de Estratégia Orçamental (DEO), disponibilizado pelo Ministério das Finanças (MF) [11].

Contudo, saliente-se que os valores finais das previsões para as diferentes variáveis foram obtidos pela média aritmética dos valores disponibilizados pelo FMI, OCDE e CE, sempre que disponíveis, em cada ano. Optou-se pelo valor disponibilizado por estas três entidades por diversos fatores. Primeiramente, por serem entidades externas ao país e estarem mais incólumes a pressões ou benefícios pela adulteração de valores; segundo, por serem as entidades que disponibilizam mais anos de previsão; e, por último, por serem as entidades que apresentam menor erro quando comparadas as médias das suas previsões com o dado histórico, no que diz respeito ao ano de 2013.

De referir que, no caso do PIB, como as previsões são fornecidas pela variação relativa ao ano anterior e o histórico expresso em valores absolutos, foi necessário proceder à transformação da variação em valor absoluto, através da equação abaixo apresentada:

$$ABS_n = ABS_{n-1} \cdot (1 + Var_n) \quad (2.1)$$

onde:

- ABS , representa o valor absoluto;
- Var , é o valor da variação;
- n , o ano em questão.

De seguida, poderão ser observados os valores referentes ao histórico e às previsões relativas ao PIB, à TInf e à TDes.

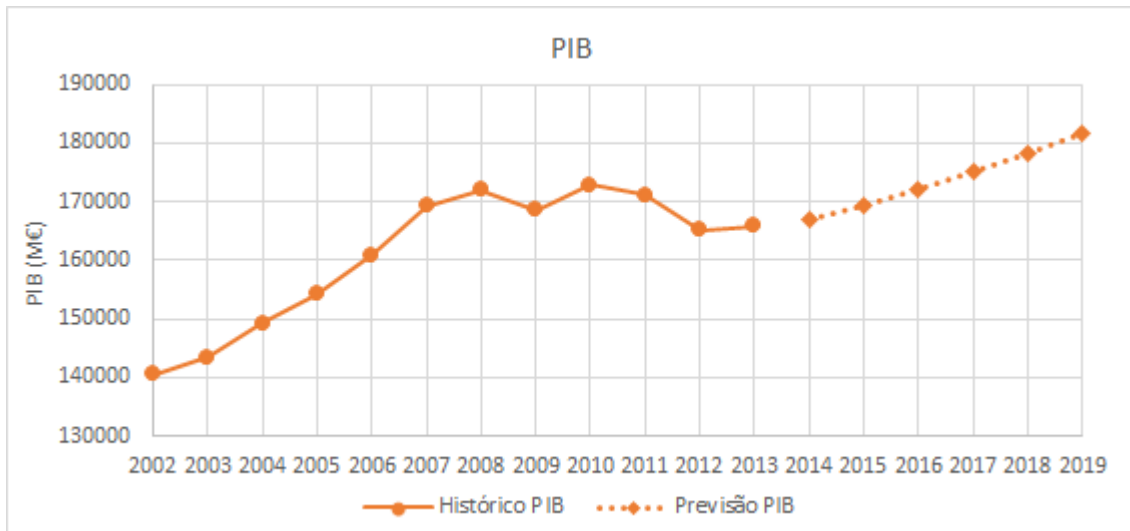


Figura 2.8 - Histórico e previsão do PIB

O PIB apresenta um comportamento muito idêntico ao que ocorre para a energia veiculada na rede, onde se observa um abrandamento depois de 2006, seguida de uma quebra e, posteriormente, com uma previsão que indica uma recuperação menos acentuada do que a existente nos primeiros anos de histórico.

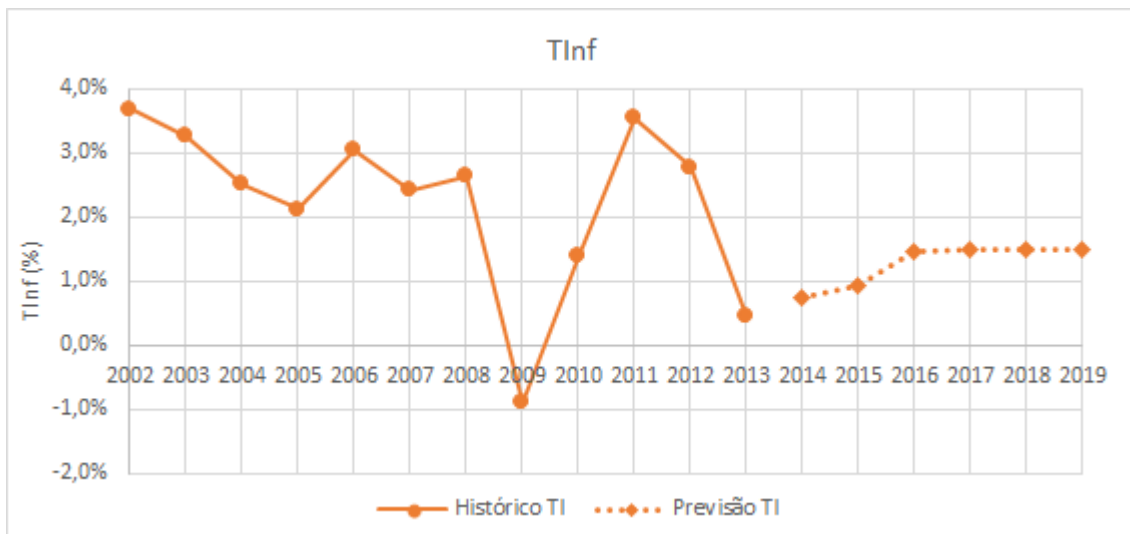


Figura 2.9 - Histórico e previsão da TInf

A TInf apresenta um comportamento oscilatório e, por isso, distinto das restantes variáveis já apresentadas, sendo que em 2009 apresenta uma deflação. Esta oscilação poderá fornecer informação relevante nos processos de estimação, como, por exemplo, delinear o início de uma crise económica, demarcada pela deflação ocorrida em 2009.

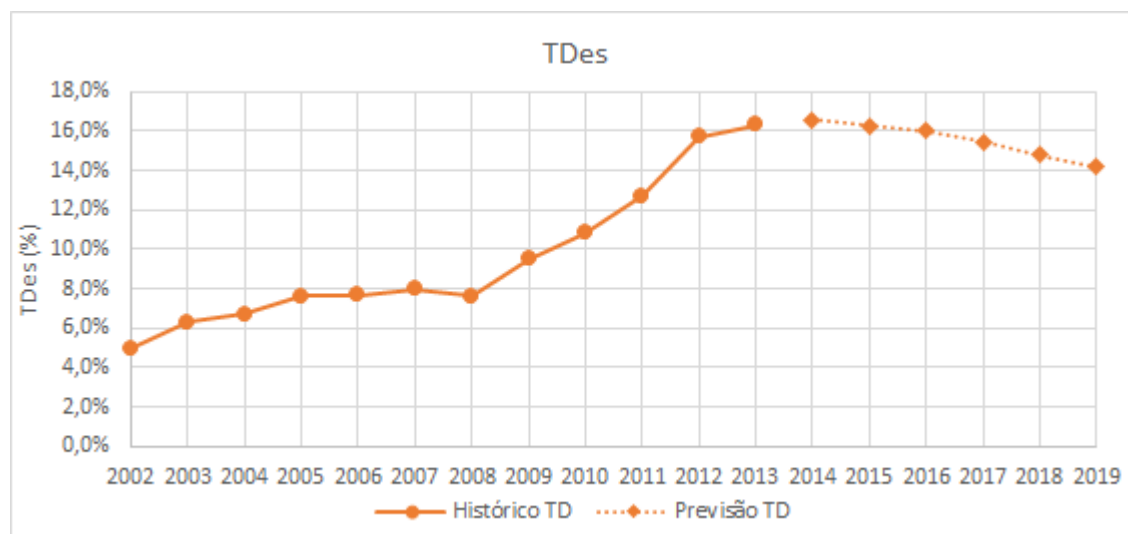


Figura 2.10 - Histórico e previsão da TDes

A TDes comporta-se de forma distinta das outras variáveis, na medida em que aumenta consoante pior é a situação socioeconómica do país. Mais uma vez, observa-se uma subida íngreme desde 2009, o que poderá representar, novamente, o início da situação de crise, estagnando entre 2012 e 2014, decrescendo ligeiramente nos anos posteriores. No entanto, devido a uma quase estagnação seguida de uma brusca variação, nos últimos anos do histórico, poderá ser difícil treinar o ME onde esta se inclua.

Indicadores relativos à Construção Civil - IC, NNF e NNC

Como referido no enquadramento, os IO são maioritariamente provenientes da ligação à RND de novos clientes, de urbanizações ou de lotes, pelo que se revela de grande importância a inclusão destes indicadores na sua estimação. Por estas mesmas razões, percebe-se que estes são também relevantes na estimação das realizações por parte da EDP Distribuição, detalhadamente, NPPT, CNCS e CNLA, posteriormente apresentadas.

O histórico foi obtido desde 2002 até 2013 para o IC e até 2012 para o NNF e NNC, através do *website* do INE.

O IC foi adquirido com periodicidade mensal, designado “índice de produção na construção e obras públicas - bruto” com base em 2005, datando de última atualização a 15 de Novembro de 2013. O IC, como descrito atrás, é um dado global, ou seja, referente à totalidade do país. Como este tem uma periodicidade mensal, procedeu-se à média aritmética dos meses de cada ano. De referir que o histórico do IC, referente ao ano de 2013, só se encontra disponível até ao mês de Setembro, pelo que ainda pode sofrer algumas alterações.

De acordo com duas notícias publicadas pela Federação Portuguesa da Indústria da Construção e Obras Públicas (FEPICOP), o setor da construção terminou 2013 com uma queda

global de 15% [12]. No entanto, 2014 iniciou com sinais de melhoria na atividade, prevendo-se que, no final do ano, a queda constante analisada na última década se reduza a uma queda de 4,5%, mostrando-se otimistas para o futuro da atividade neste setor [13].

Como o *site* do INE permite extrair o NNF e o NNC discriminados por conselho, foi possível distribuir os seus valores por DRC. Para isso, foi necessário observar os concelhos pertencentes a cada DRC, com recurso a um ficheiro Excel, disponibilizado pela EDP Distribuição, que contém os conselhos pertencentes a cada uma destas.

Foi extraído um ficheiro por cada DRC, no *website* do INE, que contém o NNF e o NNC, distribuídos pelos conselhos previamente selecionados. Depois de extraídos os ficheiros, procedeu-se ao somatório de todos os conselhos, do qual se obtém os valores distribuídos por DRC, em que, por sua vez, o seu somatório reverte os valores globais. Este procedimento pode ser observado nos respetivos ficheiros já referidos, presentes na pasta “Base para Dados”, que acompanham o documento da dissertação em formato digital.

As previsões deste índice foram obtidas por processos de estimação, descritos na secção 2.3.

Em seguida, é possível observar os dados históricos e as previsões destes indicadores.

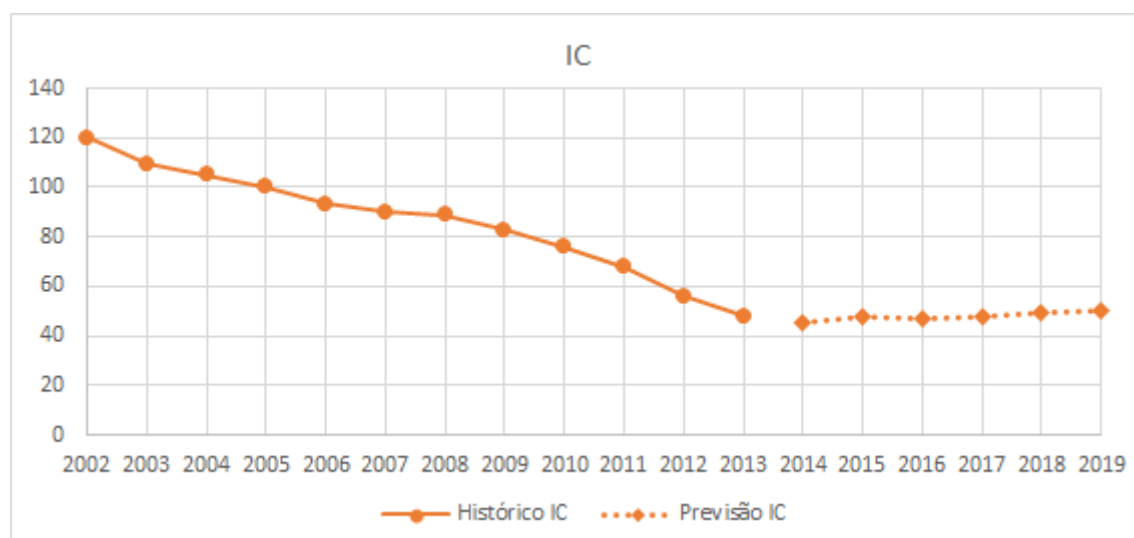


Figura 2.11 - Histórico e previsão do IC

De acordo com as previsões anunciadas pela FEPICOP, o IC apresenta um abrandamento da queda que decorria há mais de uma década no setor, seguido de uma ligeira melhoria nos próximos anos.

O NNF e o NNC foram estimados por regressões afetas ao IC, pelo que apresentam uma tendência futura análoga ao mesmo. Este processo encontra-se descrito na secção 2.3.

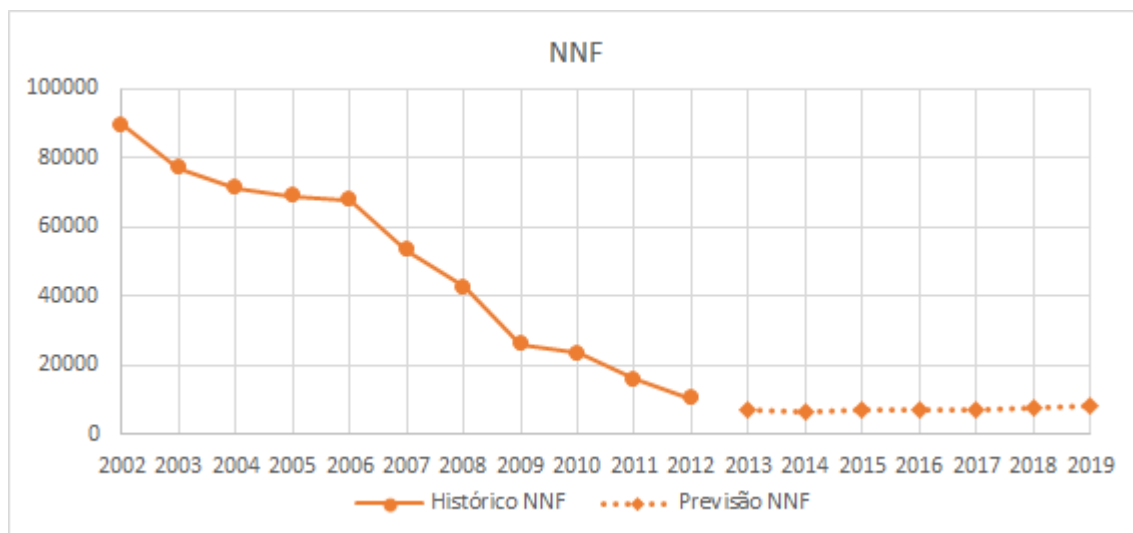


Figura 2.12 - Histórico e previsão do NNF

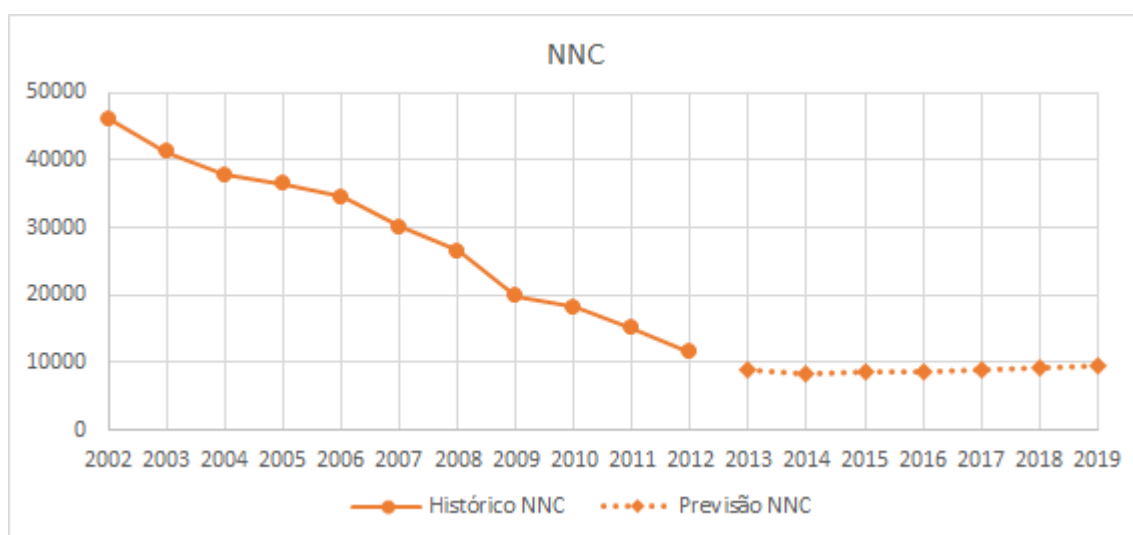


Figura 2.13 - Histórico e previsão do NNC

Realizações - NPPT, CNLA e CNCS

O termo “realizações” é usado neste âmbito para significar intervenções na rede, maioritariamente associadas ao reforço ou construção de novas linhas, postos de transformação e subestações.

O histórico relativo a estas variáveis foi fornecido pela EDP Distribuição, desde 2002 até 2011, detalhado por conselho. De forma idêntica ao NNF e NNC, para se obter os valores relativos a cada DRC, somam-se os valores dos conselhos referentes a cada uma, que por sua vez, somados, dão origem aos valores globais de cada realização, por parte da EDP distribuição.

As realizações em PT, LA e CS apresentam um histórico muito volátil e, inesperadamente, sem relação entre si. Isto implica uma atenção acrescida aquando da realização das previsões por processos de estimação, expostos na secção 2.3. É ainda importante referir que, por vezes,

parte das realizações só são contabilizadas em anos posteriores, o que pode explicar picos existentes em alguns anos do histórico das realizações, possíveis de serem observados nas figuras seguintes.

Nos gráficos seguintes apresentam-se os valores globais para as realizações.

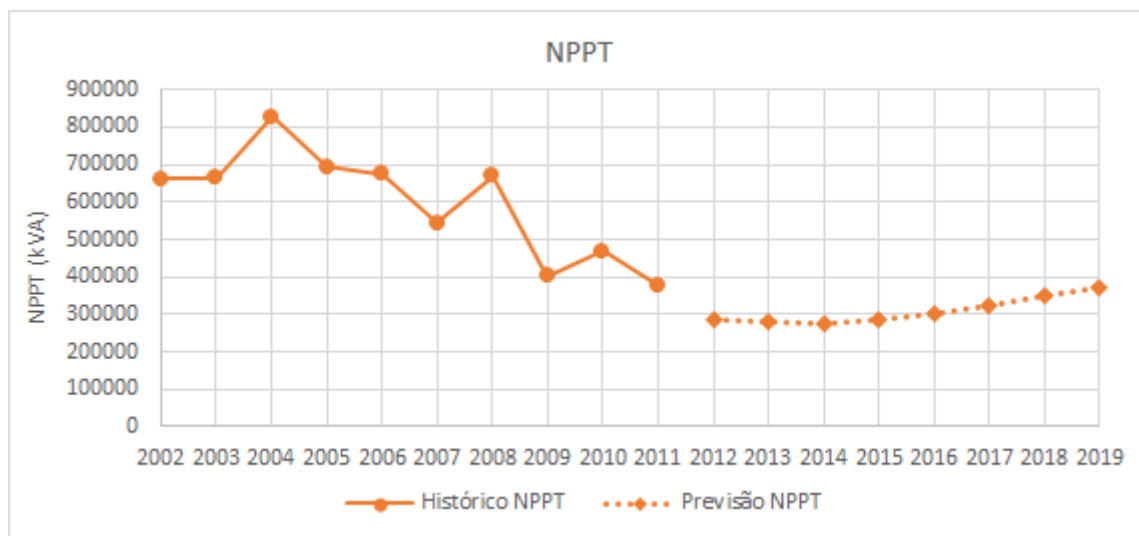


Figura 2.14 - Histórico e previsão da NPPT

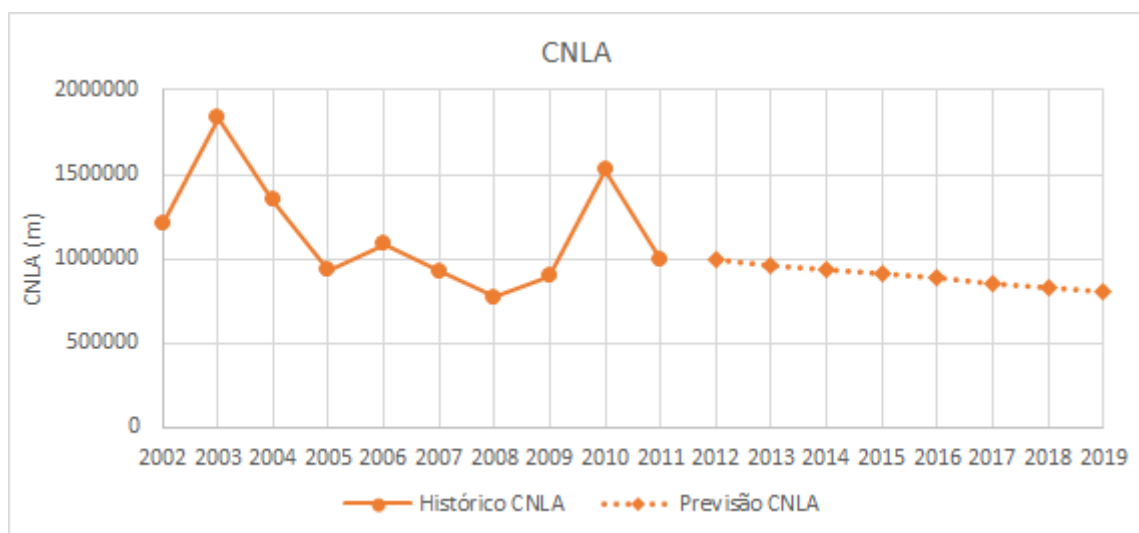


Figura 2.15 - Histórico e previsão da CNLA

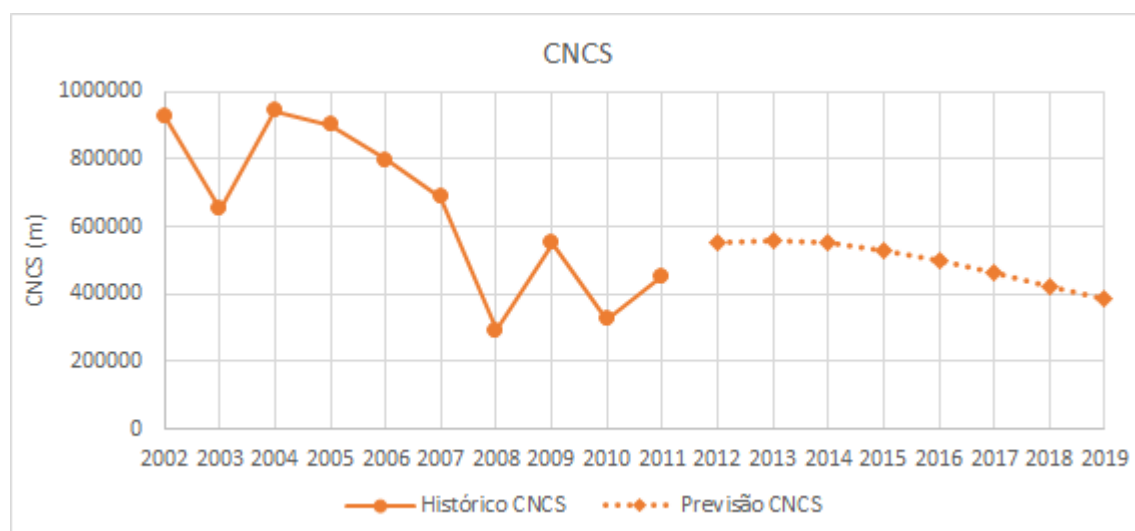


Figura 2.16 - Histórico e previsão da CNCS

2.1.2 - Alterações Especiais

Os processos de estimação consistem no treino de coeficientes com base na relação do histórico entre as entradas e as saídas. Posto isto, uma vez que são estes coeficientes a fornecer, posteriormente, as estimações, é conveniente que o conjunto de treino seja tão coerente quanto possível. Havendo situações perfeitamente excecionais identificáveis e conhecendo-se os seus impactos nos dados, faz sentido tentar corrigir estes comportamentos anómalos.

De seguida, são apresentadas as razões e a forma como se procedeu ao ajuste de valores em certas variáveis.

Ajuste das CE

Ao detetar-se uma inconsistência de valores nas CE em relação às CF, no período de 2007 a 2009, optou-se por retificar estes valores, após acordo com a EDP Distribuição. Muitas vezes, as CE, decorrentes de uma nova ligação executada num determinado ano, só são contabilizadas no ano posterior. Isto leva a uma evolução atípica das CE, pelo que se decidiu proceder à redistribuição destes valores de forma a apresentarem uma curva similar às CF.

O processo utilizado na redistribuição dos valores das CE consistiu numa distribuição proporcional à observada nas CF. Primeiramente, foi necessário dividir o valor referente a cada um dos três anos das CF pelo seu somatório, obtendo-se a percentagem que cada ano representa do total. Posteriormente, multiplicou-se esta percentagem pelo somatório das CE, obtendo-se assim o valor corrigido. A expressão utilizada no processo descrito é dada por:

$$CE'_n = \left(CF_n / \sum_{n=2007}^{2009} CF \right) \cdot \sum_{n=2007}^{2009} CE \quad (2.2)$$

onde:

- CE' , é o valor corrigido da CE;
- CE , o valor da CE antes da correção;
- CF , o valor da CF;
- n , o ano em questão.

Esta correção foi aplicada a todas as UO. No entanto, a título de exemplo, são apresentados, de seguida, os valores globais das CF e das CE, antes e depois de aplicada a correção.

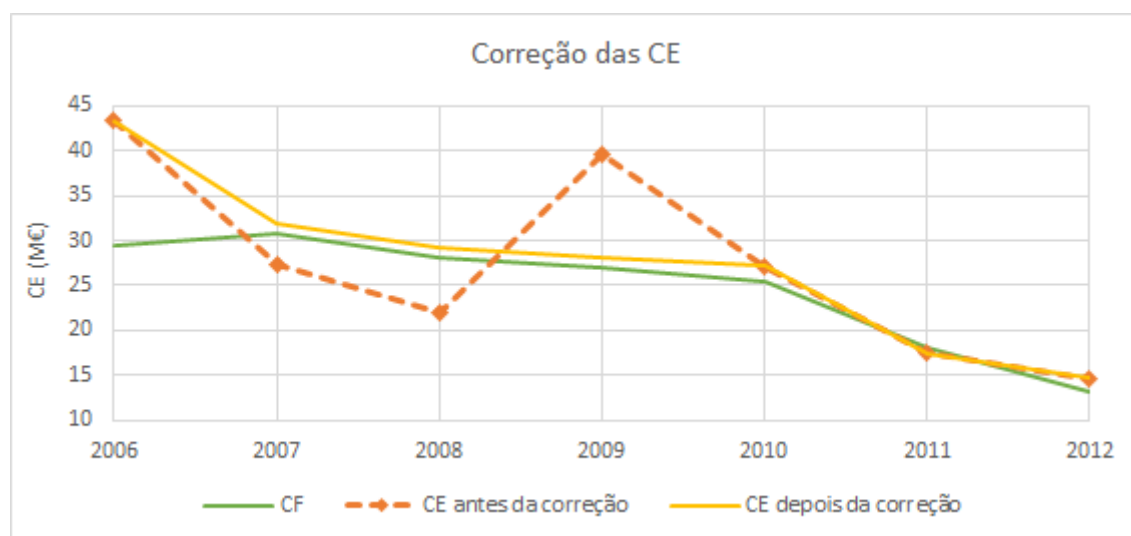


Figura 2.17 - CE antes e após correção

Serra da Serpa

O projeto para a eletrificação da Serra da Serpa foi financiado por fundos europeus. Visto que os montantes envolvidos não se relacionam com os índices macroeconómicos do país, nem da região, foi necessário ajustar os valores históricos dos IO e das participações, de forma a não incluírem os valores afetos a este projeto.

Com o consentimento da EDP Distribuição, que cedeu os montantes afetos a este projeto recebidos entre o ano de 2010 e 2013, procedeu-se à subtração destes nos valores referentes à DRC do Sul. Estes montantes, pelos motivos de confidencialidade já descritos acima, encontram-se multiplicados por um coeficiente e podem ser observados na tabela seguinte:

Tabela 2.1 - Fundos Europeus destinados à eletrificação da Serra da Serpa

Ano	Eletrificação da Serra da Serpa (€)
2010	98000
2011	659500
2012	829000
2013	635000

Note-se que os valores destas variáveis para 2013 são referentes a estimações, obtidos por ME que não contemplam no seu histórico os montantes referentes a este projeto. Torna-se, assim, desnecessário subtrair o montante afeto ao projeto para o ano de 2013 às estimações efetuadas.

De forma a perceber a dimensão do projeto e os montantes envolventes a nível global e da DRC Sul, são abaixo apresentados os gráficos que contemplam os montantes antes, nas séries representadas com (*) e representadas a cor mais escura, e após o ajuste dos IO, das CF e das CE.

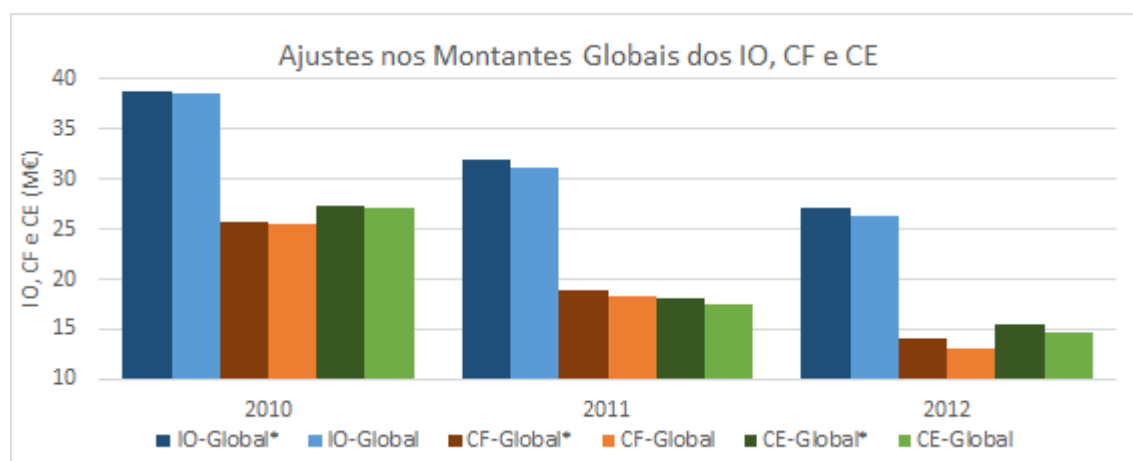


Figura 2.18 - Impacto dos ajustes nos montantes globais dos IO, CF e CE

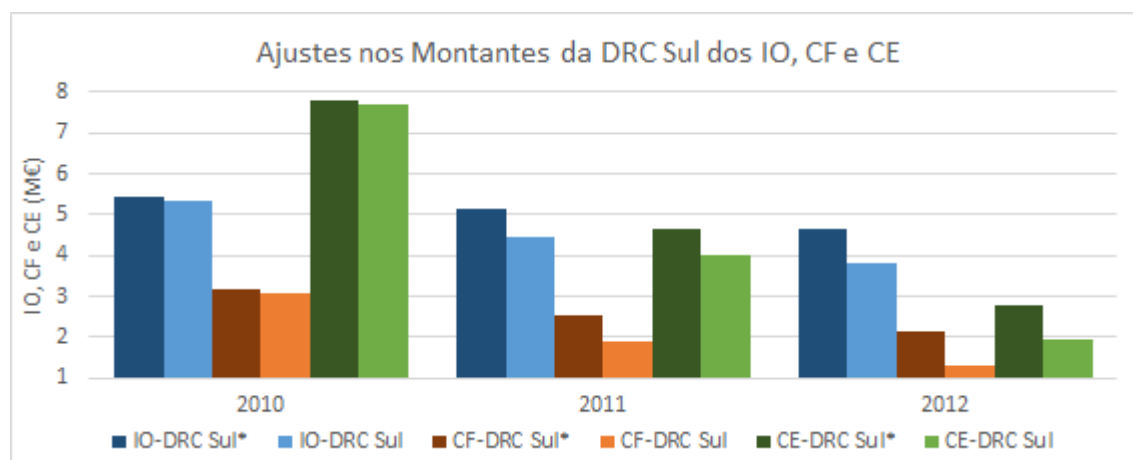


Figura 2.19 - Impacto dos ajustes nos montantes da DRC Sul dos IO, CF e CE

Nos gráficos observa-se que os ajustes têm um maior impacto nos montantes da DRC do Sul. Revela-se com isto que este ajuste não representa um impacto significativo nas estimações globais, mas sim na sua distribuição pelas DRC. Observa-se também na DRC do Sul que as CF apresentam um comportamento incomum, em comparação com os valores globais nos anos de 2010 e 2011, ao possuírem valores superiores.

2.1.3 - Impacto de Alterações na Legislação

Neste ponto será abordado o impacto que as alterações da legislação implicam na evolução dos IO, CF e CE.

Nos anos de 2007 e de 2013 entraram em vigor alterações na legislação, com impacto direto nas regras de cálculo dos montantes a suportar diretamente pela EDP Distribuição, os designados por CAPEX, e dos montantes a suportar pelos requisitantes de novas ligações, através de CF e CE. O CAPEX, que se define pelo montante efetivamente despendido pela EDP Distribuição na aquisição de equipamentos necessários ao bom funcionamento da RD, é obtido pelo IO subtraído das CF, tendo em conta o ponto da linha a considerar para efeitos comerciais de cálculo dos encargos de ligação.

Alteração na legislação em 2007

A alteração ocorrida em 2007 foi introduzida por modificações no estipulado no Capítulo VI do Regulamento de Relações comerciais (RRC) de 2005 [14], publicado a 31 de Agosto, referente a condições comerciais de ligações às redes, aprofundadas pelo despacho da ERSE nº 12 741/2007, de 21 de Junho de 2007 [15], um mês antes das alterações entrarem em vigor. A ERSE publicou ainda o despacho nº 6402/2011, a 14 de Abril [16], onde introduziu um conjunto de clarificações e alterações de menor relevância afetos ao tema.

Sucintamente, as alterações introduzidas em 2007 distinguem o ponto de ligação físico, de um cliente a uma rede, do ponto de ligação a utilizar para efeitos de cálculo comerciais dos encargos decorrentes dessa ligação, que até à data era o mesmo. O ponto de ligação físico, onde se encontram reunidas as condições técnicas para interligar a rede, é especificado pelo ORD, após requisição do pedido de ligação, enquanto o ponto de ligação para efeitos comerciais passa a corresponder ao ponto da rede, afeto ao mesmo nível de tensão solicitado na requisição, que se encontra mais próximo da instalação a ligar, independentemente de nesse ponto existirem, ou não, condições técnicas para efetuar a ligação em causa. A Figura 2.20, abaixo apresentada, ilustra esta distinção, onde o ponto de ligação físico e o ponto a considerar para o cálculo dos encargos de ligação à rede são representados, respetivamente, por A e B.

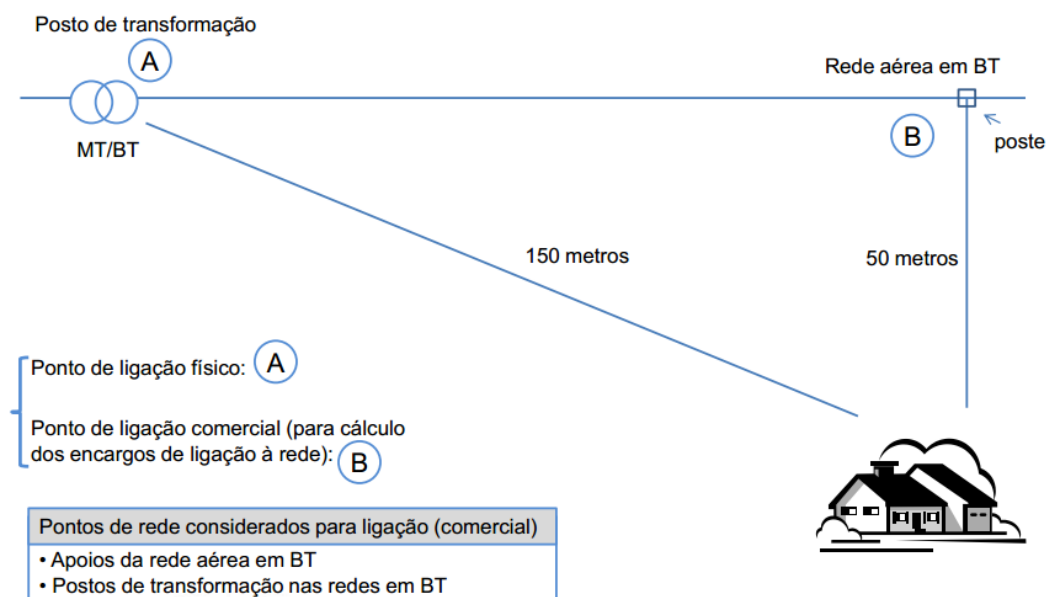


Figura 2.20 - Ilustração dos conceitos de ponto de ligação físico e ponto de ligação para efeitos comerciais (Fonte: ERSE: “Ligações às Redes de Energia Eléctrica”, Setembro de 2011, disponível em www.erse.pt)

No caso ilustrado, a ligação terá um custo efetivo correspondente ao ramal que liga ao ponto A, com uma extensão de 150 metros, sendo este o ponto que reúne as condições necessárias para a ligação requisitada (por ex. nível de tensão, potência e número de fases). No entanto, as CF pagas pelo requisitante são calculadas com base no custo de uma ligação ao ponto B, com uma extensão de 50 metros.

O RRC de 2005 estabelece também uma distinção entre elementos de uso exclusivo e elementos de uso partilhado, decorrentes da ligação. Os elementos de uso exclusivo são todos aqueles por onde é previsto transitar energia elétrica, produzida ou consumida, exclusivamente, pela instalação em causa. Já os elementos de uso partilhado são caracterizados por permitirem a ligação a mais do que uma instalação [17].

Acresce ainda que para efeitos de cálculo dos encargos da ligação à rede, os elementos de uso exclusivo têm um comprimento máximo de 30 metros para BT e 250 metros para MT, sendo que o excedente passa a ser considerado como elementos de uso partilhado, podendo o ORD sobredimensioná-los, com vista à ligação de outras instalações. Do mesmo modo, se estes forem construídos pelo requisitante e excederem o comprimento máximo, o ORD devolve ao requisitante o montante relativo à construção que exceda o comprimento máximo estabelecido para os elementos de uso exclusivo [17].

Para ligações em BT e MT, os encargos decorrentes da construção de elementos de uso exclusivo são calculados com base nos orçamentos prévios dos trabalhos a desenvolver e suportados pelo requisitante até ao limite dos encargos correspondentes ao comprimento máximo anteriormente especificado. Por sua vez, os encargos com elementos de uso partilhado são calculados com base numa expressão a utilizar para o efeito, especificada no mesmo RRC, em função da potência requisitada e da extensão dos elementos da ligação. Notar ainda que

para ligações em AT ou Muito Alta Tensão (MAT), os encargos decorrentes da ligação à rede são objeto de acordo entre o operador da rede e o requisitante da ligação [17].

No entanto, visto que estes encargos são calculados de forma análoga para todas as novas instalações, variando apenas o nível de potência, percebe-se que o custo do reforço das redes a montante, que uma nova ligação, por vezes, implica, não está contemplado nos mesmos.

Note-se ainda que apesar de menor importância no que respeita aos montantes referentes aos IO, CF e CE, a regulamentação em vigor estabelece ainda:

- o requisitante poderá construir não apenas os elementos de uso exclusivo mas igualmente os de uso partilhado, desde que haja acordo explícito do Operador da Rede;
- o Operador de Rede poderá solicitar ao requisitante a cedência de um local para instalação de um PT, devendo, neste caso, indemnizar o requisitante com um valor obtido segundo as regras estabelecidas no Regulamento de Relações Comerciais;
- uma especificação dos valores relativos à orçamentação de ligações em BT e em MT;
- após construção dos elementos de ligação, estes passam a fazer parte integrante das redes, logo que sejam considerados em condições técnicas de exploração pelo Operador da Rede [17].

As alterações até aqui descritas, que entraram em vigor em 2007, resultaram na redução dos montantes das CF e das CE devidas pelos requisitantes de novas ligações e no aumento do IO efetivamente suportado pela EDP Distribuição, designado por CAPEX. Para melhor perceção das consequências que a alteração na regulamentação teve no CAPEX, ou seja, na evolução dos valores referentes aos IO e CF, é apresentado um gráfico na Figura 2.21. Neste gráfico, observa-se, no eixo da esquerda, a percentagem que as CF e o CAPEX têm nos IO. No eixo da direita do mesmo gráfico, observam-se os valores absolutos, em M€, dos IO, das CF e do CAPEX.

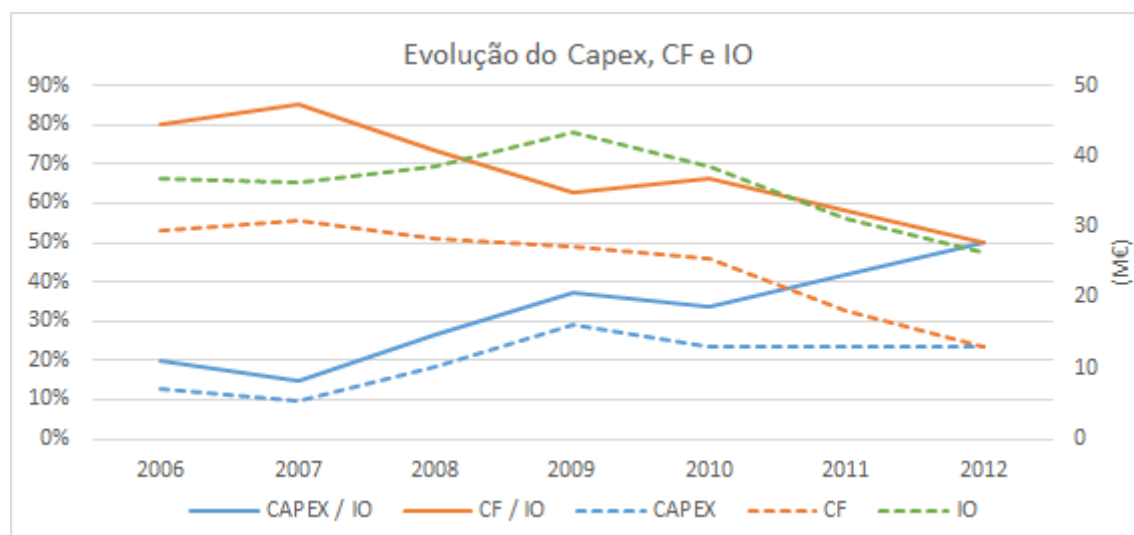


Figura 2.21 - Evolução do CAPEX das CF e dos IO decorrente da alteração na legislação, no período referente ao histórico

Como se observa no gráfico, confirma-se que após 2007 as CF têm vindo a reduzir acentuadamente, pelo que deixaram de representar cerca de 80% dos IO, passando, em 2012, a sensivelmente 50%. O inverso verifica-se em relação ao CAPEX, ou seja, a EDP Distribuição deixou de ter um custo efetivo próximo de 20% para cerca de 50% dos IO. De notar que em termos absolutos apesar do CAPEX ter estagnado nos anos posteriores a 2009, o IO esteve em contínuo decréscimo. É plausível ponderar que estas alterações possam estar também relacionadas com a atenuação do NTC, entre 2007 e 2013, conforme observado na Figura 2.7.

Porém, devido à escassez de dados históricos, é sempre difícil recolher conclusões com rigor.

Alteração na legislação em 2013

A alteração ocorrida a 12 de Maio de 2013 foi introduzida por alterações no estipulado no Capítulo X do RRC de 2012 [18], publicado a 12 de Novembro, referente a condições comerciais de ligações às redes, e pela diretiva da ERSE nº 18/2012 [19], publicada a 8 de Novembro. Estes foram aprovados por deliberação do Conselho de Administração da ERSE, a 25 de Outubro de 2012. Ao introduzir as alterações constantes desta diretiva, a ERSE pretende aumentar a eficiência de todo o processo decorrente da ligação, incluindo uma melhoria dos incentivos económicos a transmitir aos requisitantes, de modo que o custo suportado por estes aumente em função da distância a que se encontra. Ao mesmo tempo, promove uma afetação eficiente dos recursos, de tal modo que se a potência requisitada pelos requisitantes aumentar ou custos a suportar pelos mesmos também aumentem.

De forma sucinta, serão apresentados os pontos referentes a alterações relevantes face à anterior legislação, na evolução dos IO das CF e das CE, afetos aos encargos decorrentes de uma nova ligação à rede [20]:

- O ponto de ligação à rede para efeitos de cálculo dos encargos passa, novamente, a ser definido como o ponto da rede mais próximo da instalação a ligar que disponha das condições técnicas para o efeito, como ditava a legislação anterior a 2007. Estes pontos de ligação encontram-se, por exemplo, em armários de distribuição, na rede subterrânea em BT, ou apoios de rede na rede aérea em MT[20].

- Como na antiga legislação, para BT, continua a existir diferenciação na totalidade da ligação dos elementos para uso exclusivo em relação aos de uso partilhado. O comprimento máximo dos elementos para uso exclusivo manteve-se em 30 metros, de forma ao excedente ser considerado como elementos de uso partilhado, ou seja, idêntico ao que se verificava. No entanto, para MT passou a considerar-se somente elementos de ligação para uso partilhado.

- Nas ligações em MAT, AT e MT só as que excedem os 2 MVA é que continuam com os encargos decorrentes da ligação à rede, a ser objeto de acordo entre o operador da rede e o requisitante. Para as ligações em BT e MT inferiores a 2 MVA, os encargos com a ligação são calculados de acordo com as regras estabelecidas pela ERSE [20].

Como consequência destas alterações, antevê-se um aumento das CF e das CE face aos IO, contrariamente ao que sucedia com a alteração ocorrida em 2007. No entanto, prevê-se um aumento menos acentuado, mas possivelmente mais acelerado.

De referir também que as alterações ocorridas não se refletiram de forma imediata, mas sim ao longo de sensivelmente dois a três anos.

Contudo, é possível modelizar os efeitos decorrentes destas alterações recorrendo a sigmóides, do tipo decrescente para as de 2007 e crescente para as de 2013, típicas de uma resposta sobrearmortecida de um sistema a uma alteração brusca. Com a junção destas sigmóides, é possível obter uma curva que se adapte às alterações ocorridas, com efeitos nos montantes das comparticipações, designada por curva de adaptação.

As sigmóides, decrescente e crescente, que dão origem à curva de adaptação são definidas pelas equações (2.3) e (2.4), respetivamente:

$$VS_n = L_{m\acute{a}x} \cdot e^{-\left(\frac{Ano_n - Ano_0}{\tau}\right)^2} + VS_{Inicial} \quad (2.3)$$

$$VS_n = L_{m\acute{a}x} \cdot k \cdot \left(1 - e^{-\left(\frac{Ano_n - Ano_0}{\tau}\right)^2}\right) \quad (2.4)$$

onde:

- VS_n , é o valor devolvido pela sigmoide no ano n ;
- $L_{m\acute{a}x} = VS_{Inicial} - VS_{Final}$, e representa a amplitude máxima de variação da sigmoide;
- Ano_n , o ano para o qual se quer obter o valor da sigmoide;
- Ano_0 , ano em que se inicia o declive;
- τ , tempo em anos para a sigmoide atingir 63,21% da variação máxima;
- k , constante em %, que permite moldar a amplitude máxima de variação da sigmoide crescente em relação á decrescente.

A título de exemplo e para valores padrão, são apresentadas na Figura 2.22, a curva de adaptação e as sigmóides que lhe deram origem. Neste exemplo, considerou-se que o valor máximo e mínimo que a sigmoide toma é de 1 e 0, respetivamente. Os anos em que se inicia o declive para a sigmoide, crescente e decrescente, são afetos aos das alterações, isto é, 2007 e 2013, respetivamente. O valor considerado para o coeficiente τ foi de 3, ou seja, três anos para atingir 63,21%, de $L_{m\acute{a}x}$ no caso da sigmoide decrescente e de $L_{m\acute{a}x} \cdot k$ no caso da crescente. A constante k , que neste exemplo toma o valor de 90%, surgiu com o propósito da EDP Distribuição ter indicado a antevisão de que as comparticipações, apesar de muito próximas, não voltariam a representar uma percentagem tão elevada dos IO como as que ocorriam nos anos antecedentes à primeira alteração na legislação, em 2007.

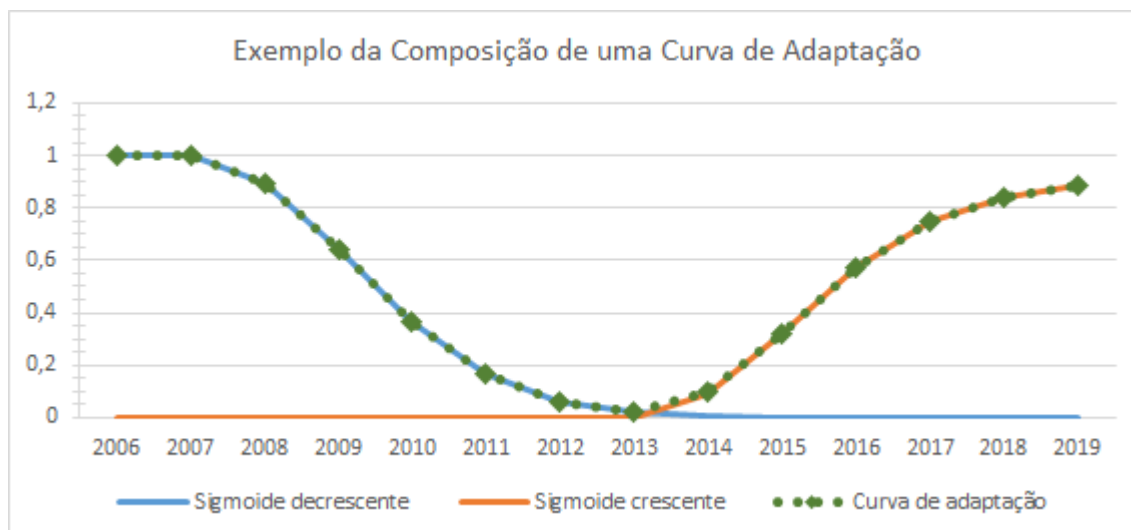


Figura 2.22 - Exemplo da composição de uma curva de adaptação, através da composição de sigmóides que modelizam o impacto de alterações na legislação

Efetivamente, estes conceitos são essenciais aquando da realização das estimações das participações. Este tema voltará a ser abordado e aplicado na subsecção 2.3.4, referente à estimação destas variáveis.

2.2 - Pré-processamento de Dados

Esta secção tem em vista complementar o anterior tratamento de dados, de modo a otimizar os recursos disponíveis. Ou seja, pretende-se retirar o máximo partido da informação afeta às variáveis e, ao mesmo tempo, testar ME alternativos, aumentando, assim, a possibilidade de obter melhores resultados aquando das estimações. Além dos ME que contemplam os dados absolutos passam a existir mais três, com dados normalizados, com as variações e com a dessensibilização.

Um processo de estimação treina os coeficientes com base nas entradas, com o intuito de melhor explicarem a saída. Com isto, torna-se necessário criar ME para realizar os processos de estimação, ou seja, organizar conjuntos de variáveis para verificar qual destes é o que melhor se aproxima das saídas.

A informação é fornecida aos processos de estimação através dos valores afetos a cada variável. Para retirar o máximo partido dessa informação, é necessário aplicar diversas transformações aos valores. Estas encontram-se apresentadas nesta secção, sendo que a primeira consiste numa normalização dos dados, a segunda na variação destes em relação ao ano anterior e, por último, a dessensibilização dos mesmos.

Normalização de dados

As variáveis disponíveis para criar os ME possuem dimensões de escala muito diferentes umas das outras. Este facto é redundante no processo de cálculo das estimações com recurso a regressões, na medida em que o resultado final da estimação advém da operação dos coeficientes com os dados de entrada. No entanto, no processo de otimização dos coeficientes, ou seja, do treino destes, não o é, na medida em que o processo usado para determinar esses coeficientes terá maior facilidade em fazê-lo se não estiver a trabalhar com entradas de dimensões muito díspares.

Contudo, a abordagem escolhida para a standardização de dados foi o método Min-Max [7], que permite ao utilizador ajustar os dados com base no máximo e mínimo histórico, para uma gama de valores por ele definida e assim melhorar a eficiência do treino da rede.

De notar que foram experimentados outros tipos de normalização. No entanto, este é o que melhor se adapta às exigências de contradomínio, dada a dificuldade de alguns processos de regressão no tratamento de variáveis negativas. Como exemplo temos a regressão logarítmica que não admite como entradas valores negativos ou 0. Tipicamente o intervalo definido para esta normalização é de [-1 a 1]. No entanto de forma a não eliminar o peso dos coeficientes, optou-se por definir um intervalo entre [1 a 3] e assim eliminar valores próximos de zero.

A equação usada por este método na standardização de valores encontra-se abaixo representada:

$$v' = \frac{v - \min_a}{\max_a - \min_a} \cdot (\max_A - \min_A) + \min_A \quad (2.5)$$

onde:

- v' , representa o valor normalizado;
- v , o valor absoluto;
- a , é referente aos valores absolutos da variável;
- A , é referente ao valor da escala em que se pretende mapear os valores.

Esta equação permite mapear os valores absolutos de uma variável, compreendidos na escala entre o mínimo e o máximo dos seus valores, definida pelo intervalo $[\min_a, \max_a]$, em valores compreendidos numa escala definida pelo utilizador, representada pelo intervalo $[\min_A, \max_A]$ [7].

Na Tabela 2.2 é apresentado um exemplo com o intuito de visualizar o efeito decorrente desta aplicação em duas variáveis com dimensões de escala tão díspares como o PIB e a TDes.

Tabela 2.2 - Exemplo da aplicação do método Min-Máx às variáveis PIB e TDes

Ano	Valores Absolutos		Valores Normalizados	
	PIB (M€)	TDes (%)	PIB	TDes
2002	140567	5,0%	1	1
2003	143472	6,3%	1,141654	1,22478386
2004	149313	6,7%	1,426472	1,29394813
2005	154269	7,6%	1,668155	1,44956772
2006	160855	7,7%	1,989346	1,46685879
2007	169319	8,0%	2,402072	1,51873199
2008	171983	7,6%	2,531973	1,44956772
2009	168529	9,5%	2,363549	1,77809798
2010	172860	10,8%	2,57471	2,00288184
2011	171126	12,7%	2,490188	2,3314121
2012	165108	15,7%	2,196694	2,85014409
2013	165854	16,3%	2,233081	2,95389049
2014	166959	16,6%	2,286999	3
2015	169241	16,3%	2,398267	2,9481268
2016	172118	16,0%	2,538565	2,90201729
2017	175216	15,5%	2,689641	2,80691643
2018	178370	14,8%	2,843436	2,6945245
2019	181581	14,2%	3	2,5907781
min_a	140567	5%	min_A	1
max_a	181581	17%	max_A	3

Variação anual dos dados

Esta abordagem pretende obter a estimação das saídas com base na variação das entradas (diferença entre o valor num ano relativamente ao ano anterior). Por vezes, o valor referente à variação dos dados permite aos processos de estimação realizarem estimações mais próximas das saídas. Estes passam a treinar os coeficientes com base em ME constituídos pela variação dos valores afetos a cada variável e não pelos valores absolutos.

De notar que, devido à necessidade de possuir um dado relativo ao ano anterior, no cálculo da variação, é eliminado um dado referente ao primeiro ano histórico em cada variável.

De seguida, é apresentada a equação que permite calcular a variação dos dados em relação ao valor do ano anterior para cada variável:

$$Var_n = \frac{Abs_n}{Abs_n - 1} - 1 \quad (2.6)$$

onde:

- Var , representa a variação;
- Abs , o valor absoluto;
- n , o ano em questão.

Como exemplo, na Figura 2.23 abaixo apresentada, visualiza-se no eixo da esquerda o valor absoluto e no da direita o valor referente às variações, em percentagem, para a TDes.

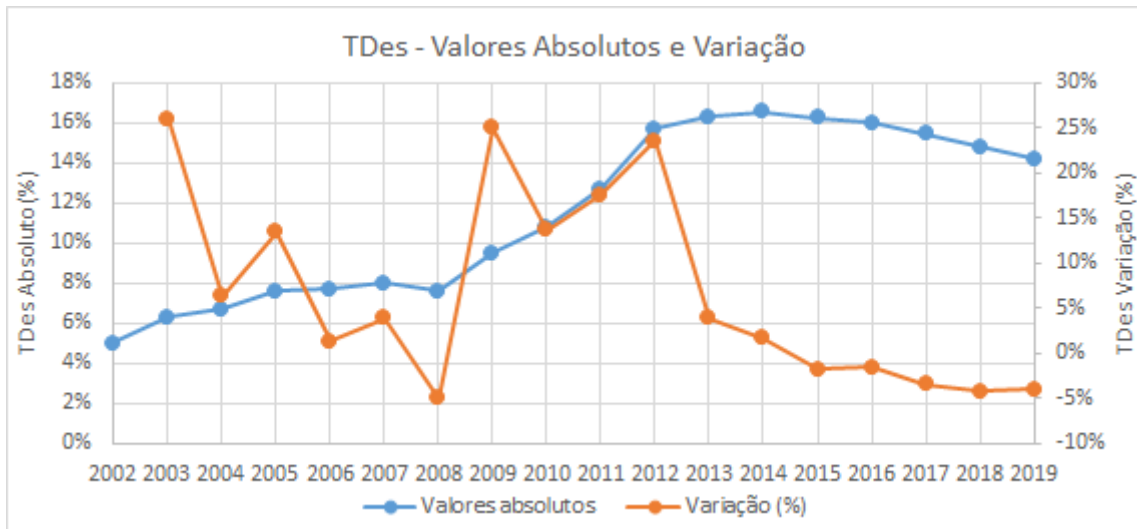


Figura 2.23 - Ilustração da aplicação do método da variação anual dos dados na variável TDes

Dessensibilização das entradas

Ao dessensibilizar as entradas pretende-se suavizar a importância dos valores afetos às variáveis de entrada, tornando a saída menos dependente (menos sensível) às variações dessas entradas. A aplicação desta técnica é baseada no pressuposto que algumas variáveis explicativas apresentam ocasionalmente oscilações inexplicáveis, ou pouco consistentes, com a saída ou com outras entradas. Consegue-se, com isto, suavizar possíveis erros, suscetíveis de estarem contidos em alguns dos anos de uma certa variável requerida no treino dos coeficientes.

Na sua versão mais simples, esta técnica leva a que cada exemplo (entradas, saída) é replicado duas vezes, o que resulta em 3 padrões:

- o original;
- a entrada a ser dessensibilizada é afetada negativamente (decrece um pouco);
- a entrada a ser dessensibilizada é afetada positivamente (aumenta um pouco);

O aumento e a redução são determinados por um coeficiente, designado coeficiente de dessensibilização. Este coeficiente é expresso em percentagem e, neste trabalho, considerou-se ser metade do coeficiente de variação de uma variável. Optou-se por uma margem de apenas metade do coeficiente, de forma a não suavizar demasiado os valores e com isso comprometer a informação disponível. O coeficiente de variação, designado em estatística por coeficiente de variação de Pearson, é dado pelo quociente entre o desvio padrão e a média.

Determinado o coeficiente de dessensibilização e tendo por base o valor real (central), calcula-se o valor acima e abaixo deste, através da respetiva equação:

$$V' = V \cdot (1 \pm C_D) \quad (2.7)$$

onde:

- V' , representa os valores dessensibilizados;
- V , o valor real;
- C_D , o coeficiente de dessensibilização.

A título de exemplo, é abaixo apresentada uma tabela que contempla a aplicação da dessensibilização à variável PIB, para o período de 2002 a 2005:

Tabela 2.3 - Exemplo da aplicação da dessensibilização à variável PIB

Ano	PIB (M€)	Ano	PIB (M€)
2002	140566,8	+	145315,1
2003	143471,7	2002	140566,8
2004	149312,5	-	135818,5
2005	154268,7	+	148318,1
...	...	2003	143471,7
		-	138625,3
		+	154356,2
Média	165374,4	2004	149312,5
Desvio Padrão	11172,5	-	144268,8
		+	159479,8
C _{Varição}	6,76%	2005	154268,7
C _{Dessensibilização}	3,38%	-	149057,6
	

Abaixo, na Figura 2.24, é visível o efeito da aplicação da dessensibilização à variável PIB, onde são apresentados os valores do PIB antes e após a dessensibilização.

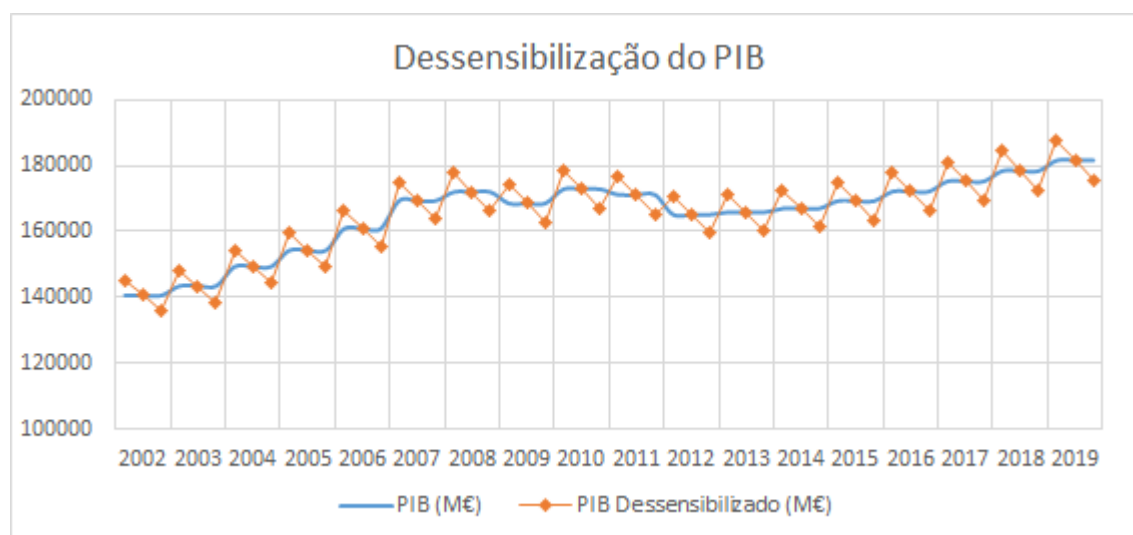


Figura 2.24 - Ilustração da aplicação da dessensibilização nos dados da variável PIB

Com aplicação da dessensibilização nas entradas, como referido, é possível criar novos exemplos, os quais, embora não contenham nova informação, ajudam a controlar os eventuais

problemas de sobreajuste aos dados. Na medida em que os coeficientes são treinados, em função da mesma saída, não só para o valor original mas também para os afetados, negativa e positivamente.

Na Figura 2.25, são apresentadas como exemplo três estimações, referentes a ME que contém as mesmas variáveis, mas para diferentes abordagens, isto é, valores de base, variação e dessensibilização dos dados.

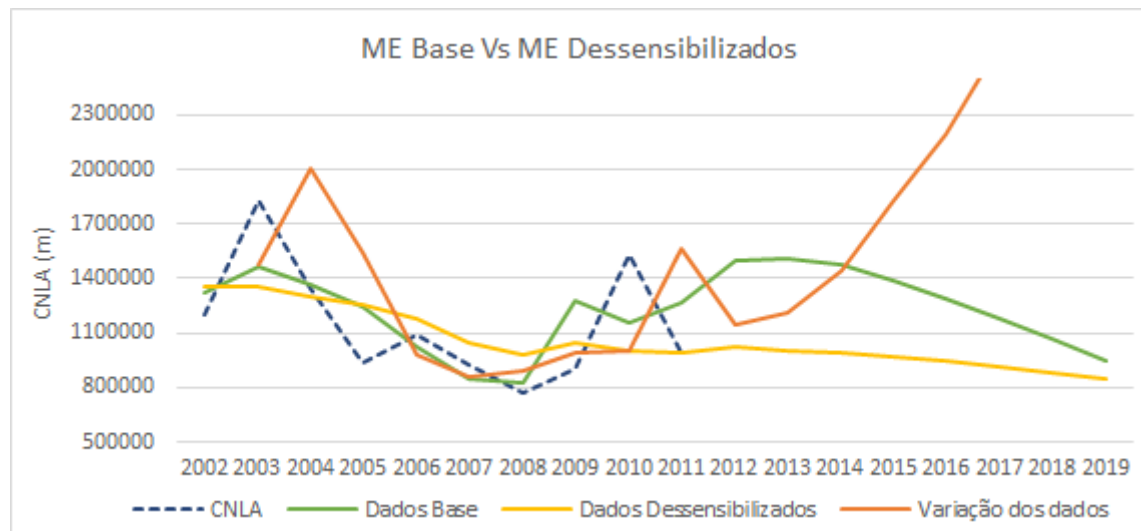


Figura 2.25 - Estimativas para a CNLA obtidas pelos ME(EnMTBT, PIB, NNF) para dados de Base, Variação e Dessensibilizados

Optou-se pela escolha desta variável para exemplo, pelo fato dos ME apresentarem dificuldade em se aproximarem dos valores históricos. Isto deve-se, possivelmente, ao facto desta apresentar, nos anos de 2003 e 2010, picos para os quais não se arranja explicação nem se encontra grande relação com as restantes variáveis. No entanto, como se observa, a dessensibilização das entradas permitiu eliminar a influência desses picos, aquando do treino dos coeficientes.

No ME que usa como entradas a variação dos dados, é perceptível uma das limitações desta abordagem. Como o treino dos coeficientes é feito com base na variação, esta abordagem é suscetível de fornecer estimativas com picos muito elevados, ou seja, com um desvio padrão elevado. De notar que o ME que usa a variação dos dados foi descartado por um processo descrito na secção 2.4.

2.3 - Processos de Estimação

Como descrito anteriormente, é necessário realizar não só as estimativas dos objetivos principais do trabalho, isto é, dos IO, CF e CE, mas também das variáveis para as quais ainda

não existem previsões, referidas na subsecção 2.1.1, possíveis de conter informação benéfica à estimação.

Estas variáveis são referentes ao setor da construção civil, tais como o IC, o NNF e o NNC e às realizações por parte da EDP Distribuição, nomeadamente, o NPPT, CNLA e CNCS.

Na estimação destas, recorreu-se às variáveis que continham todos os dados, tanto históricos como de previsões, com o intuito de observar quais as que constituíam os ME que melhor se adaptavam às mesmas.

No entanto, é necessário observar a ordem com que se irá realizar a estimação das previsões em falta, de forma acrescentar mais informação aos modelos nas estimações posteriores, pois algumas destas necessitam de mais informação do que outras para serem estimadas. Com isto, é necessário observar como as variáveis em falta se relacionam com as restantes variáveis já disponíveis, abaixo apresentado. Esta análise é feita na subsecção 2.3.1.

Nos pontos seguintes deste subcapítulo, serão expostos os processos que deram origem às estimações de cada variável, bem como de outros testes realizados, de forma a conhecer quais os processos e regressões que melhor se adaptam.

Os ME experimentados foram constituídos por duas, três ou quatro das variáveis disponíveis. Com isto cria-se um imenso número de ME, dado pela combinação sem repetição das variáveis disponíveis agrupadas em conjuntos de duas, três ou quatro. Por exemplo para seis variáveis disponíveis, número disponível para realizar a primeira estimação, consegue-se obter $C_{(6,2)} + C_{(6,3)} + C_{(6,4)}$, isto é, cinquenta ME. Este número, que aumenta com o número de variáveis disponíveis, quando multiplicados pelas três abordagens aos dados, nomeadamente, valores originais, variação e dessensibilização, geram uma combinação de ME, impossíveis de serem diagnosticados na íntegra.

Para dar resposta a este entrave, optou-se por desenvolver um programa em Matlab que otimiza todo o processo de estimação. Este, descrito sucintamente, consiste em gerar ME para todas as combinações de variáveis disponíveis, realizar as estimações, diagnosticar os modelos, ordenar por ordem de mérito os modelos de acordo com a sua avaliação, e por último filtrar os modelos que apresentam estimações fora dos padrões. Este encontra-se detalhado na secção 2.4, abaixo apresentado.

Como forma de avaliar os melhores ME decidiu-se descartar do histórico disponível o último ano para ser usado como diagnóstico do ME e ficar com todos os restantes para treino. Optou-se por este critério mais uma vez devido ao baixo histórico existente e com isto disponibilizar ao treino de coeficientes o máximo de informação. A razão para o ano disponível para teste ser o ano mais próximo das estimações, deve-se ao facto deste permitir avaliar com maior rigor se os coeficientes estão a fornecer estimações com uma tendência próxima da real, no ano mais próximo das estimações a realizar.

Pelo mesmo fator foi realizada uma média ponderada das medidas de erro, entre o erro de treino, com 40%, e o erro de teste, com 60%. As medidas de erro utilizadas como para

diagnóstico de desempenho dos ME foram o *Mean Absolute Percentage Error* (MAPE) e uma normalização do *Root Mean Square Error* (RMSE), designada pela sigla (NRMSE), que consiste na divisão do RMSE pela média do valor real. Estas medidas de erro foram calculadas respetivamente através das equações abaixo apresentadas, para os valores de treino e teste, para posteriormente ser calculada a média ponderada referida anteriormente [21].

$$MAPE = \frac{1}{n} \cdot \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \quad (2.8)$$

$$NRMSE = \frac{\sqrt{\frac{1}{n} \cdot \sum_{t=1}^n (Y_t - \hat{Y}_t)^2}}{\bar{Y}} \quad (2.9)$$

onde:

- Y_t , representa o valor real no ano t ;
- \hat{Y}_t , o valor estimado no ano t ;
- n , o número de anos a avaliar;
- \bar{Y} , a média dos valores nos n anos.

De notar que estas medidas de erro constituem apenas um indicador dos melhores ME, pelo que é necessário posteriormente analisar as curvas afetas às estimações, de forma a verificar se as previsões dadas por estas estão de acordo com os padrões de cada variável. Por vezes os ME adaptam-se bem no período de treino e teste e são diagnosticados com erros baixos, mas no entanto as previsões dadas por estes são absurdas. Como exemplo temos o IC, que como se encontra em contínuo decréscimo, pode em alguns casos nas suas estimações atingir valores negativos, o que é impraticável nesta variável.

2.3.1 - Relação dos Indicadores de Construção e das Realizações com as restantes variáveis

A análise de relação entre variáveis foi realizada com recurso a gráficos de dispersão [23] e à função Excel “PEARSON” [22]. Esta última devolve o valor do coeficiente de correlação momentânea do produto Pearson, r , que pode variar no intervalo de $[-1$ a $1]$, indicando uma maior relação¹ quanto mais próximo dos extremos se encontrar.

Aos gráficos de dispersão, adicionou-se uma linha de tendência da qual se retira R^2 . Notar que, para linhas de tendência lineares, este representa o coeficiente de correlação de Pearson ao quadrado [23]. Este, varia no intervalo de $[0$ a $1]$ e representa maior relação quanto mais próximo for de 1. Por sua vez, o coeficiente de correlação de Pearson, r , varia no intervalo de

¹ Este indicador mede a intensidade da relação linear entre as variáveis em questão.

[-1 a 1], indicando uma maior relação quanto mais próximo dos extremos se encontrar. No entanto, nos processos de estimação utilizados, os coeficientes podem assumir valores negativos, pelo que é indiferente saber em que sentido se dá a relação. Com isto, decidiu-se analisar apenas o coeficiente de determinação R^2 , expresso pela equação [21]:

$$R^2 = \frac{\text{Desvios Explicados}}{\text{Desvios Totais}} = \frac{\sum_{i=0}^{P_i} (\hat{Z}_{Pi} - \bar{Z})^2}{\sum_{i=0}^{P_i} (Z_{Pi} - \bar{Z})^2} \quad (2.10)$$

onde:

- \hat{Z}_{Pi} , é o valor dado pela linha de tendência no ponto P_i ;
- \bar{Z} , a média de todos os pontos da série;
- Z_{Pi} , o valor da série no ponto P_i .

Na Figura 2.26, é apresentada uma ilustração de um excerto de uma série, de forma a ser mais perceptível o que representa o R^2 e os respetivos desvios explicados e totais:

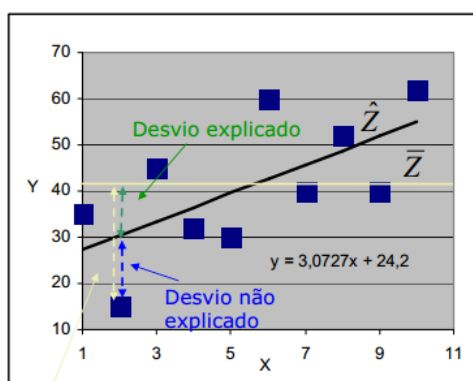


Figura 2.26 - Ilustração de um excerto de uma série com explicação dos valores afetos ao cálculo de R^2 (Fonte: ERSE: “Ligações às Redes de Energia Eléctrica”, Setembro de 2011, disponível em www.erse.pt)

Na Figura 2.27, é apresentado um gráfico de barras que indica o valor da relação, R^2 , entre as variáveis a estimar e o banco de variáveis disponíveis para constituir os ME:

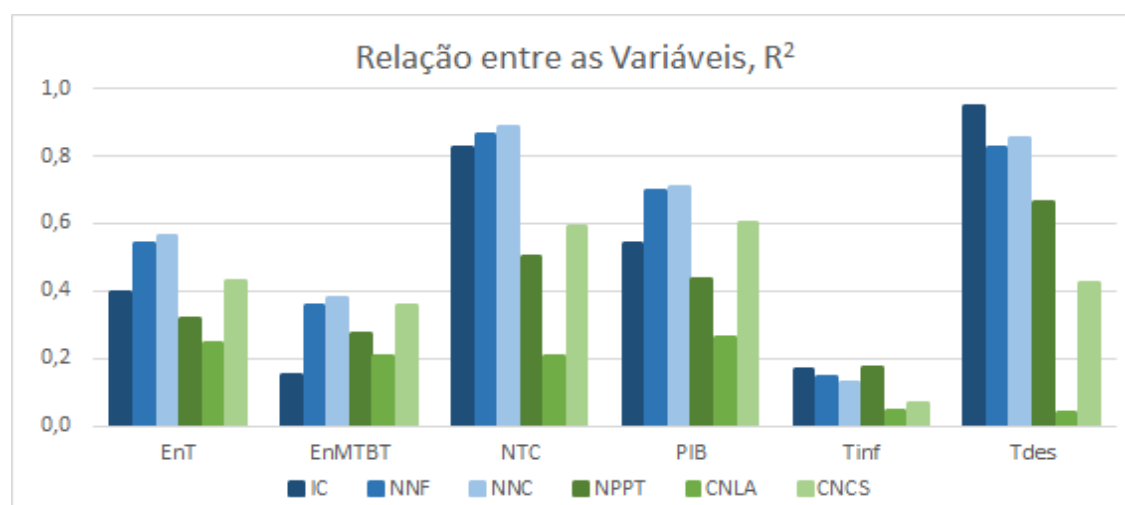


Figura 2.27 - Relação, dada por R^2 , entre as variáveis afetas ao setor da construção e às realizações, com as restantes variáveis disponíveis

O gráfico mostra que, possivelmente, os indicadores relativos ao setor da construção, representados a azul, apresentam maior relação com as restantes variáveis. No entanto, é importante referir que esta relação, por vezes, não é suficiente para caracterizar completamente a realidade. Outro aspeto a salientar é o facto do coeficiente de correlação apenas detetar relações lineares, pelo que as consequências a tirar desta análise devem ser conservativas: se o coeficiente é elevado, então existe uma relação entre as variáveis em questão; se é baixo, não é possível tirar conclusões, as variáveis podem ser relacionadas por funções não-lineares, podem ser relacionadas em conjunto com outras variáveis² ou podem não estar relacionadas.

Devido ao reduzido histórico disponível e à constante tendência de algumas variáveis, surge a possibilidade de existirem elevadas relações “falsas”, na medida em que um indicador em contínuo crescimento possua uma boa relação com outro em contínuo défice, o que não indica que quando um inverte a sua tendência o outro tenha de seguir essa mesma tendência. Isto observa-se, por exemplo, entre os indicadores de construção e o NTC, que se encontram com tendência inversa, ainda que isso não signifique que se o IC inverter a sua tendência, o NTC o faça obrigatoriamente.

Paralelamente aos indicadores de construção apresentarem uma relação superior com as restantes variáveis, estes possuem também mais dados históricos, pelo que se decidiu optar por realizar, primeiramente, as estimações dos indicadores de construção e, posteriormente, inclui-los nos modelos para a estimação das realizações.

De notar ainda que já tinham sido experimentadas outras abordagens para obter a relação entre variáveis, como o *Multidimensional Scaling* (MDS) [24] [25]. No entanto, a interpretação dos resultados mostrou-se mais complexa, possivelmente devido, mais uma vez, ao reduzido histórico existente.

2.3.2 - Indicadores de Construção - IC, NNF e NNC

Como exposto na apresentação de variáveis, o IC possui mais um ano de histórico do que as restantes variáveis e ainda um ano de previsão emitido pela FEPIOP. Isto permite fornecer aos ME mais dois anos para serem usados como treino no seu processo de estimação, pelo que será a primeira variável a ser estimada.

Devido às variáveis possuírem um histórico tão reduzido, foi necessário experimentar diversas abordagens, começando pelas mais simples até se obterem estimações que se enquadram nos padrões e com erros relativamente baixos.

² Por exemplo, quando a saída depende do produto de duas ou mais entradas.

O primeiro foi com recurso à ferramenta *solver* do Excel, a qual permite determinar os coeficientes com base num critério minimização (por exemplo, do erro quadrático). Com esta ferramenta foram testadas diversas abordagens, tais como: a inclusão nos ME de uma variável que contém os valores do IC atrasados um ano, ou seja, é testado como entrada o valor referente ao IC do ano anterior; outros tipos de regressões além da linear, nomeadamente, a exponencial, logarítmica, polinomial de segundo grau e potencial.

Contudo, verificou-se que a inclusão dos dados relativos ao ano anterior não conduziu a estimações com a qualidade desejada. Pelo contrário, verificou-se que, ao serem incluídos, estes dados tendem a reforçar excessivamente a tendência do histórico e, para um histórico tão reduzido, é imprudente realizar estimações apenas baseadas nesta tendência.

Com isto, dá-se mais importância ao coeficiente afeto a esta variável e retira-se a informação presente nas outras, realizando assim previsões com uma tendência que, muitas vezes, não se enquadra na realidade. A título de exemplo, pode ser observada, na Figura 2.28, a curva real, representada a azul, e as curvas de estimação para quatro modelos do IC. Sendo que os ME representados com linha a tracejado contêm as mesmas variáveis que os representados com linha contínua, mais a variável que contém os valores do ano anterior do IC, passando o modelo a conter mais uma variável. Estes ME foram criados com variáveis normalizadas, de forma a perceber a influência do coeficiente afeto à variável referente aos anos anteriores do IC. Estes coeficientes, bem como a NRMSE, para treino e teste, e a sua média ponderada, podem ser observadas na Tabela 2.4.

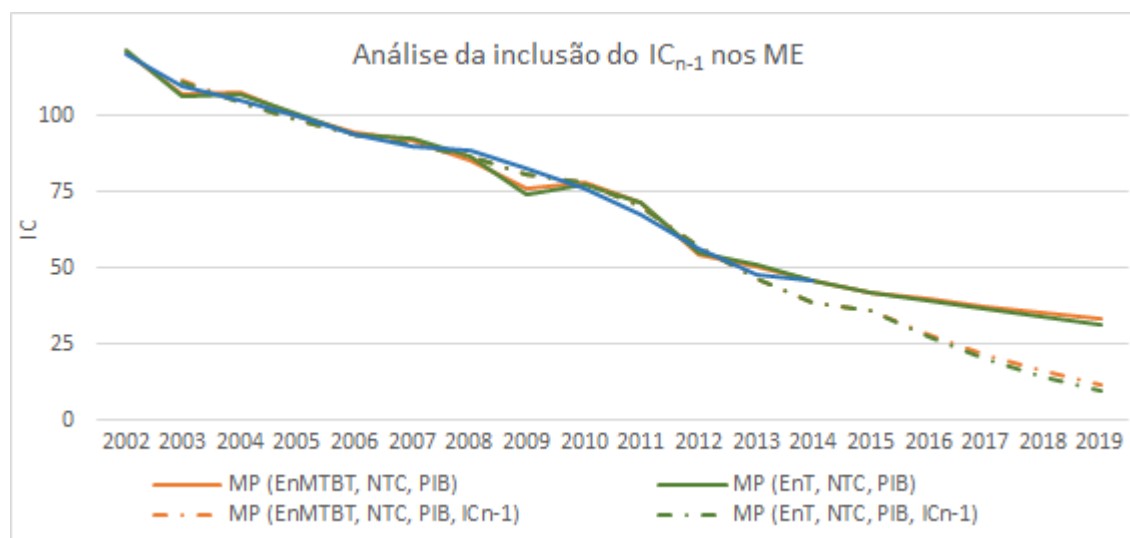


Figura 2.28 - Análise da inclusão de valores históricos de uma variável na sua previsão

Como pressuposto, as estimações que contemplam nos seus ME os valores passados do IC possuem uma tendência continuamente decrescente, análoga ao seu passado histórico.

Tabela 2.4 - Análise da inclusão de valores históricos de uma variável na sua previsão

	Coeficientes					NRMSE		Média Ponderada
	θ_0	θ_1	θ_2	θ_3	θ_4	Treino	Teste	
ME (EnMTBT, NTC, PIB)	164,1	11,5	-77,7	22,6		3,31%	0,03%	1,34%
ME (EnMTBT, NTC, PIB, ICn-1)	-22,1	1,8	-30,2	16,2	58,0	1,76%	15,59%	10,06%
ME (EnT, NTC, PIB)	166,5	16,5	-86,4	24,8		3,80%	0,83%	2,02%
ME (EnT, NTC, PIB, ICn-1)	-27,7	3,2	-29,6	15,2	59,7	1,75%	15,99%	10,30%

De referir que o coeficiente θ_0 é um coeficiente independente e que os restantes se referem, por ordem, às variáveis contidas nos ME.

Na Tabela 2.4 verifica-se que os coeficientes afetos aos valores passados do IC, que se encontram a vermelho, absorvem a importância das restantes variáveis, ao valorizarem sobretudo a adaptação do ME aos anos de treino. Observa-se assim que o erro de treino diminui, contrariamente ao que acontece para o erro de teste. Acima, na Figura 2.28, é possível visualizar o sucedido, na medida em que para 2009, os ME representados a tracejado se aproximam mais do real, enquanto para 2014, ano de teste, os ME que melhor se aproximam do real são os representados a traço contínuo.

É também de fácil perceção que se forem experimentadas todas as hipóteses de regressões disponíveis para todos os ME possíveis de criar, teremos inúmeras opções impraticáveis de serem realizadas e observadas na sua totalidade. Com isto, decidiu-se testar um ME para as diferentes regressões referenciadas, com recurso à ferramenta *Solver* do Excel. O ME escolhido contempla o conjunto de três variáveis suscetíveis de melhor se relacionarem com o IC, isto é, o NTC, o PIB e a TDes. A relação entre variáveis foi obtida com recurso a R^2 , apresentado acima, na Figura 2.27.

As estimações realizadas, com recurso às diferentes regressões, encontram-se ilustradas na Figura 2.29. Segue-se, na Figura 2.30, um gráfico de barras que ilustra os valores da NRMSE, para treino e teste, obtidos nas respetivas estimações:

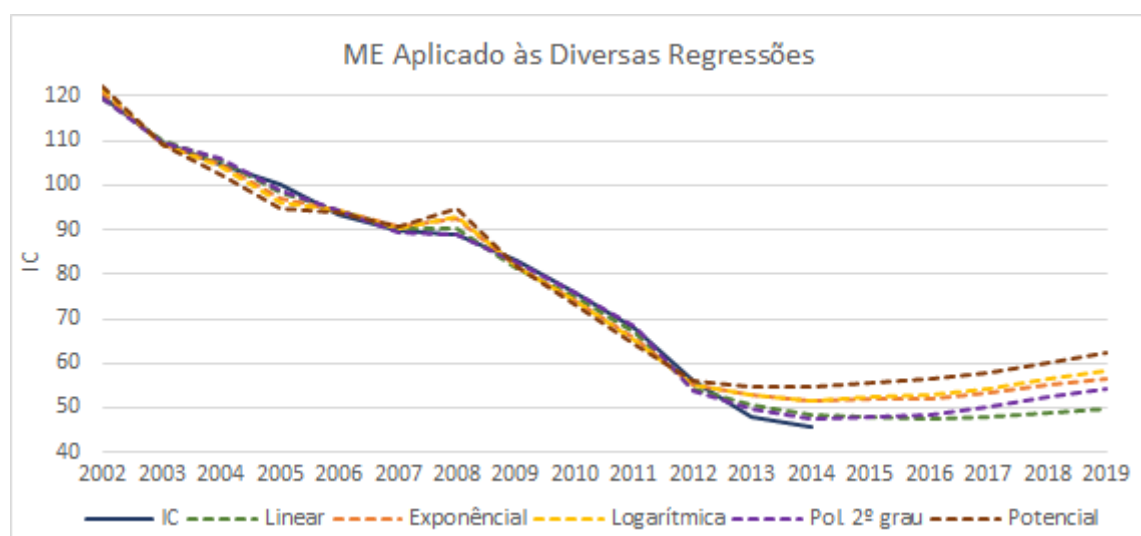


Figura 2.29 - Estimações resultantes da aplicação do ME (NTC, PIB, TDes) às diversas regressões

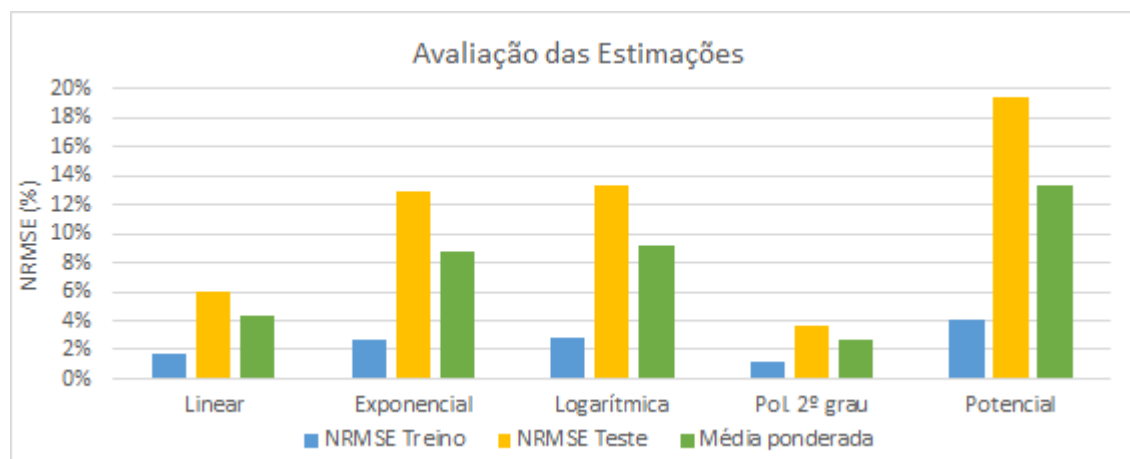


Figura 2.30 - Avaliação das estimações resultantes da aplicação do ME (NTC, PIB, TDes) às diversas regressões

Como se pode observar, a regressão linear apresenta uma boa adaptação à curva real, com resultados superiores às restantes, à exceção da polinomial de 2º grau. No entanto, esta última exige um maior número de parâmetros do modelo, o que pode prejudicar a generalização. Assim sendo, optou-se por usar uma regressão linear, devido à sua simplicidade, ao seu bom desempenho e à menor exigência de parâmetros.

Este critério tem como vantagem permitir a determinação dos coeficientes (parâmetros do modelo) sem recorrer ao *Solver*, e assim facilitar a realização e análise dum maior número de ME, pois com o *Solver* é necessário otimizar os coeficientes sempre que se introduza um novo ME, o que requer mais tempo.

Contudo, verificou-se que mesmo sem recorrer ao *Solver* do Excel, é impraticável realizar e avaliar as estimações de todos os modelos possíveis de criar. Posto isto, decidiu-se automatizar o processo de construção de ME através de programação em Matlab – o programa encontra-se descrito na secção 2.4.

Estes programas passaram a ser usados para determinar todos os ME quer das grandezas auxiliares quer das grandezas principais (IO, CF e CE). As estimações finais encontram-se expostas no Capítulo 3.

Ao aplicar a mesma metodologia à estimação do NNF e do NNC, mesmo com a inclusão do IC nos seus modelos, não se obteve sucesso. A título de exemplo, na Figura 2.31 e Figura 2.32 referentes, respetivamente, ao NNF e ao NNC, é possível observar as estimações obtidas para os três melhores modelos de três variáveis, usados na estimação do IC, e para os mesmos acrescidos da variável IC, possuindo assim quatro variáveis de entrada. De notar que a série cor de laranja é referente aos valores reais.

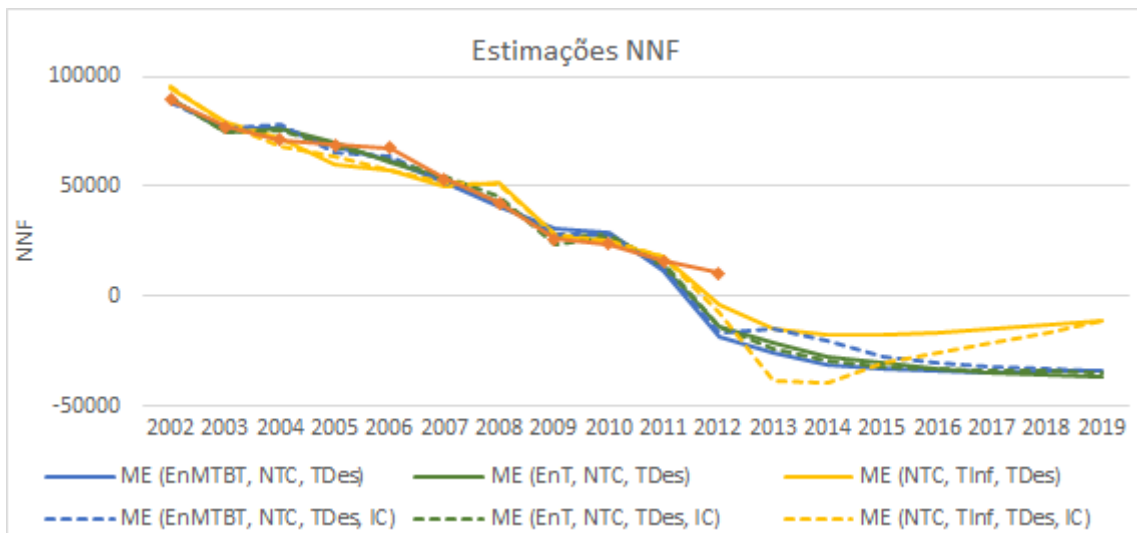


Figura 2.31 - Ilustração exemplo da aplicação do método dos mínimos quadrados na estimação do NNF.

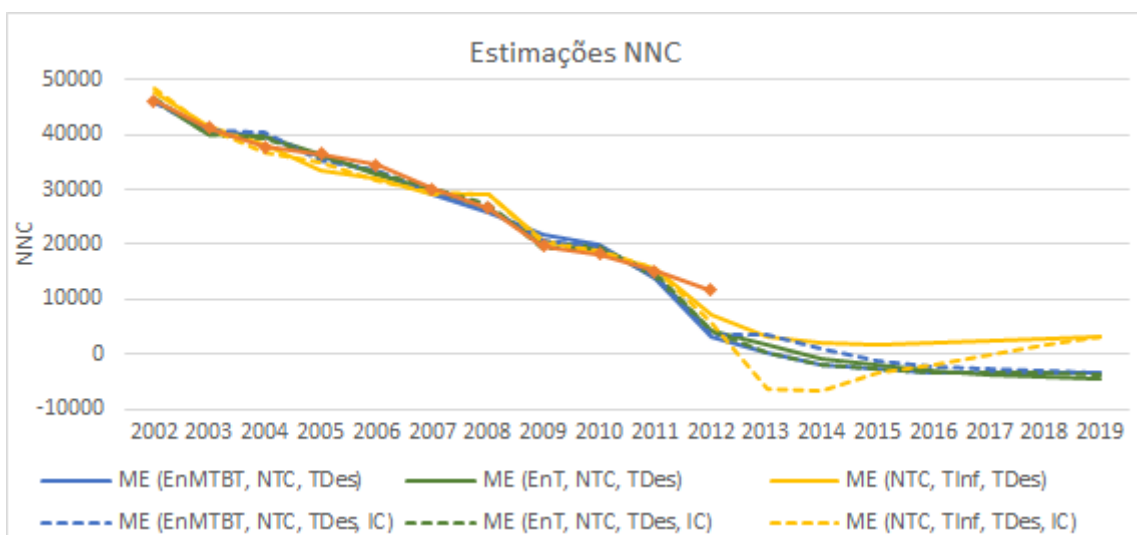


Figura 2.32 - Ilustração exemplo da aplicação do método dos mínimos quadrados na estimação do NNC.

Como se observa nos gráficos, os modelos geram previsões com uma tendência muito pessimista. Constata-se ainda que o único ME aceitável, não necessariamente com um erro baixo, é o que se encontra representado a amarelo contínuo, na estimação do NNC, pelo facto de não apresentar estimações negativas.

Na Tabela 2.5, observa-se o erro calculado para cada ME, onde é possível observar que o erro de teste, contrariamente ao erro de treino, é muito elevado.

Tabela 2.5 - Erros dos modelos usados na estimação, com recurso ao método dos mínimos quadrados, do NNF e NNC

		NRMSE		Média
		Treino	Teste	Ponderada
NNF	ME (EnMTBT, NTC, TDes)	7,22%	269,21%	164,41%
	ME (EnT, NTC, TDes)	6,10%	229,61%	140,21%
	ME (NTC, TInf, TDes)	10,37%	134,58%	84,89%
	ME (EnMTBT, NTC, TDes, IC)	6,52%	257,51%	157,12%
	ME (EnT, NTC, TDes, IC)	5,31%	226,68%	138,13%
	ME (NTC, TInf, TDes, IC)	9,84%	167,37%	104,36%
NNC	ME (EnMTBT, NTC, TDes)	4,25%	73,37%	45,72%
	ME (EnT, NTC, TDes)	3,49%	63,70%	39,61%
	ME (NTC, TInf, TDes)	5,51%	37,23%	24,54%
	ME (EnMTBT, NTC, TDes, IC)	3,98%	70,39%	43,82%
	ME (EnT, NTC, TDes, IC)	2,91%	63,52%	39,28%
	ME (NTC, TInf, TDes, IC)	5,02%	49,20%	31,53%

Contudo, como o NNF e o NNC apresentam uma relação quase direta com o IC, é possível que a sua estimação seja guiada unicamente pelo IC, através de regressões, nomeadamente, linear, exponencial, logarítmica, polinomial de segundo grau e potencial. Os coeficientes usados pelas diversas regressões, na estimação do NNF e do NNC, foram extraídos das equações referentes a cada linha de tendência, dos respetivos gráficos de dispersão. Abaixo podem ser observados os gráficos de dispersão, onde no eixo do X se encontra o IC e no do Y, o NNF ou NNC, respetivamente, bem como a equação de onde se extraiu o valor dos coeficientes e o valor da relação entre variáveis, fornecido por R^2 .

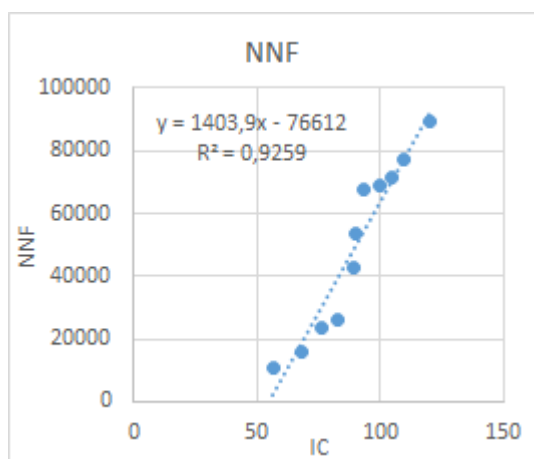


Figura 2.33 - Relação linear entre o NNF e o IC.

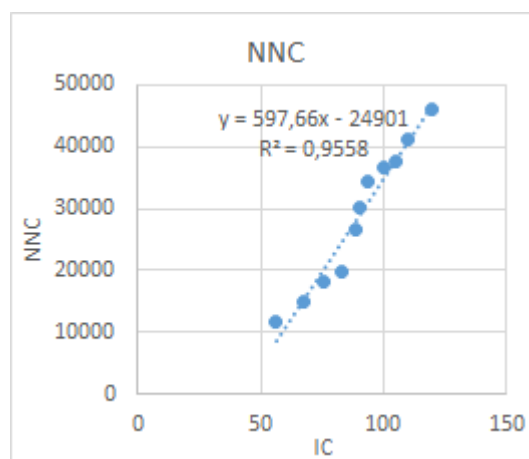


Figura 2.34 - Relação linear entre o NNC e o IC.

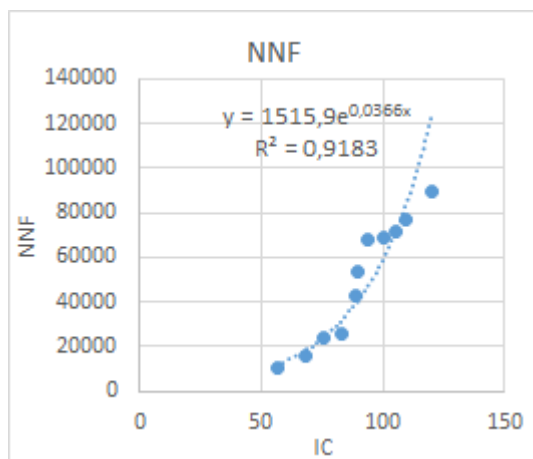


Figura 2.35 - Relação exponencial entre o NNF e o IC.

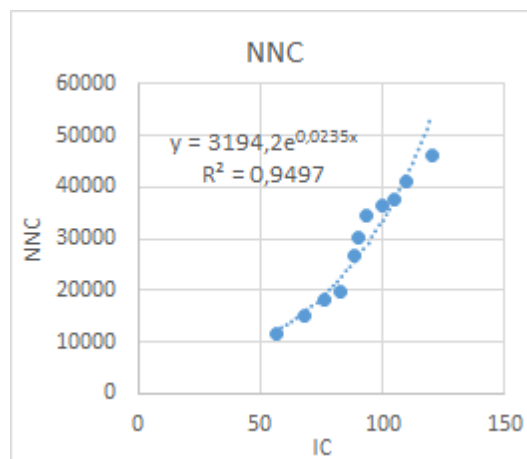


Figura 2.36 - Relação exponencial entre o NNC e o IC.

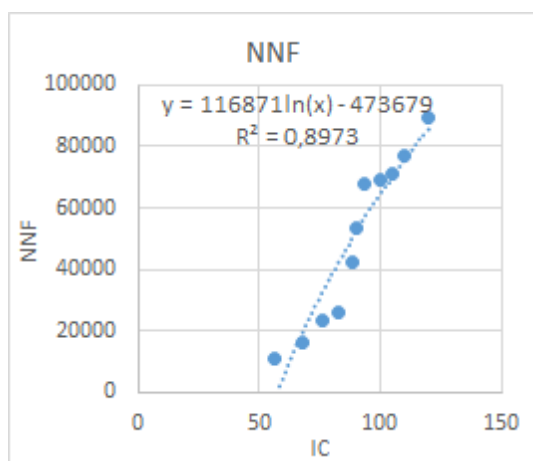


Figura 2.37 - Relação logarítmica entre o NNF e o IC.

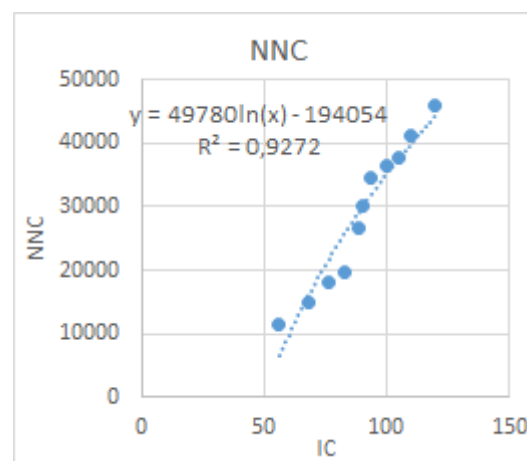


Figura 2.38 - Relação logarítmica entre o NNC e o IC.

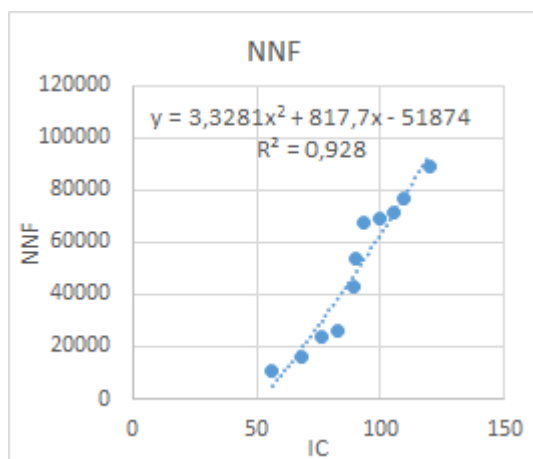


Figura 2.39 - Relação polinomial de 2º grau entre o NNF e o IC.

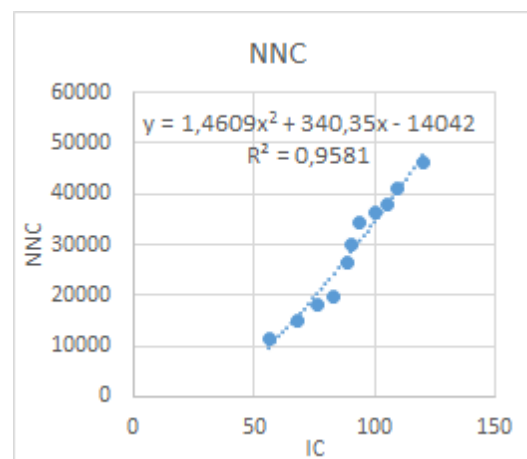


Figura 2.40 - Relação polinomial de 2º grau entre o NNC e o IC.

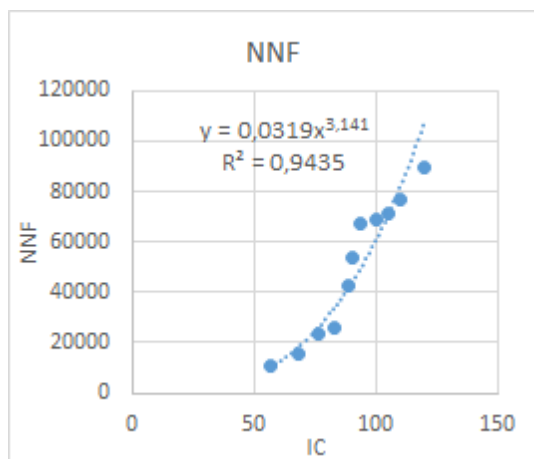


Figura 2.41 - Relação potencial entre o NNF e o IC.

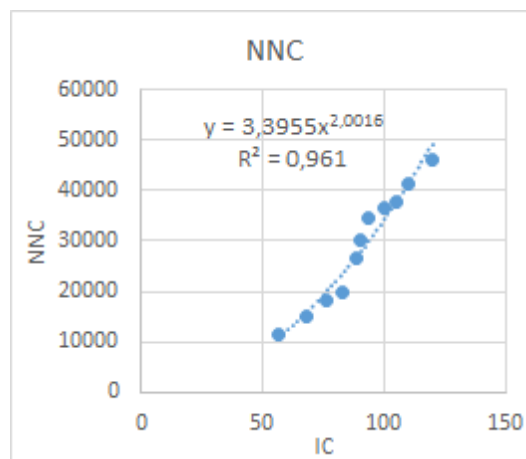


Figura 2.42 - Relação potencial entre o NNC e o IC.

Efetivamente, existe uma excelente relação (R^2 próximo de 1) do NNF e do NNC com o IC para as diversas regressões, pelo que se optou por testar estimações com base nestas variáveis.

Das estimações obtidas descartam-se as que não se encontram dentro dos padrões viáveis, como por exemplo valores negativos. Nesta fase ordenam-se por ordem de mérito as estimações obtidas pelas restantes regressões. Em seguida, calculam-se as médias das N estimações obtidas, em que N varia de dois até ao número de estimações viáveis.

Deste leque de hipóteses, incluindo as estimações obtidas pelas regressões individualmente, escolhe-se a que apresentar menor erro, obtido pela média ponderada, de treino e teste, da NRMSE, calculado pela equação (2.9). Os resultados destas estimações, bem como os respetivos erros, são expostos no Capítulo 3.

2.3.3 - Realizações - NPPT, CNLA e CNCS

As realizações por parte da EDP Distribuição advêm da necessidade de ligação de clientes e do aumento de potência que transita nas redes, de forma a manter os níveis de qualidade da energia e segurança nas mesmas. Como referido na subsecção 2.1.1, estas descrevem as intervenções na rede, maioritariamente associadas ao reforço ou expansão da rede.

Estas variáveis apresentam uma constante volubilidade ao longo do seu histórico, sem relação entre si, como referido também na subsecção 2.1.1, podendo visualizar-se esse efeito nos anos históricos, presentes na Figura 2.43.

Tal como explicado na análise do IC, na secção anterior, devido ao reduzido histórico, não faz sentido realizar as estimações com base nos seus valores passados. Com isto, optou-se por uma metodologia alternativa à adotada anteriormente, que consiste em assumir que os valores futuros apresentarão uma variação semelhante à ocorrida no passado [25]. Na Figura 2.43, é possível observar-se o resultado da aplicação desta abordagem, também experimentada neste trabalho.

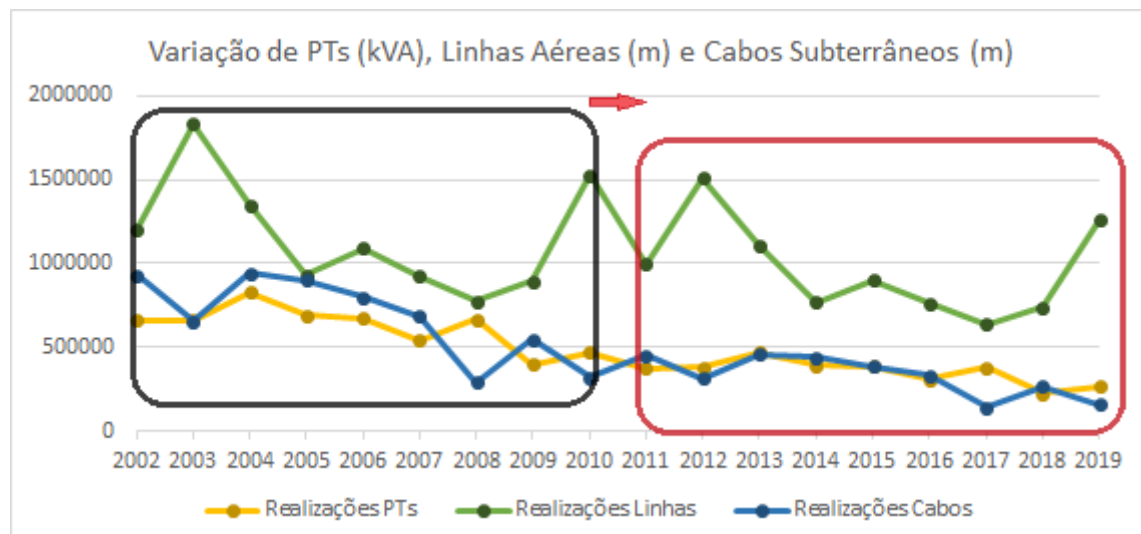


Figura 2.43 - Estimação das realizações com base na tendência de variação do histórico.

Como se pode observar, as previsões seguem exatamente a mesma tendência do histórico, delineadas, respetivamente, pelo retângulo vermelho e preto. No entanto, considerou-se que a replicação daquelas variações apenas seria aceitável se fosse possível compreender o histórico. A análise efetuada não revelou relações evidentes entre esta evolução e a das outras variáveis principais (por exemplo, IO) ou outros fatores externos (por exemplo, eleições). Assim, decidiu-se experimentar a alternativa que se descreve a seguir.

Na Figura 2.44, é apresentada a relação linear entre as realizações e os indicadores relativos à construção civil, obtida pelo coeficiente de Pearson ao quadrado, isto é, R^2 , já apresentado anteriormente. Note-se que a relação destas com as restantes variáveis disponíveis encontram-se expostas na subsecção 2.3.1, através da Figura 2.27.

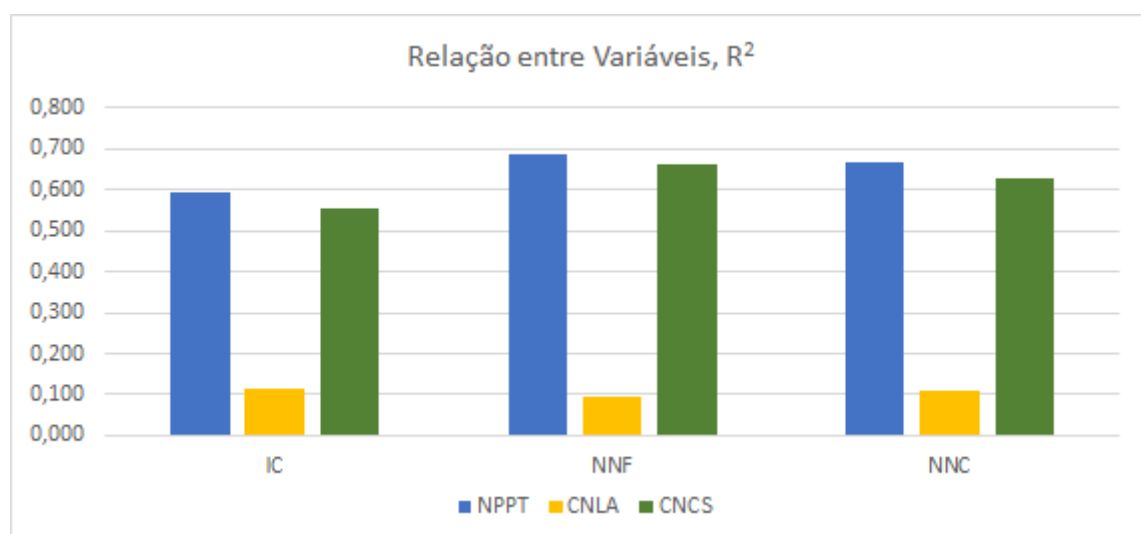


Figura 2.44 - Relação linear, dada por R^2 , entre as variáveis afetas às realizações e os indicadores relativos à construção civil.

Como se observa, existe uma relação aceitável entre as realizações e os indicadores de construção, com a exceção da CNLA que apresenta uma relação muito baixa, semelhante ao que acontece com as restantes variáveis disponíveis. Após analisar todas as linhas de tendência possíveis de aplicar aos gráficos de dispersão da CNLA em função das restantes variáveis, não se verificaram melhorias relevantes no valor de R^2 . No entanto, a composição de variáveis poderá revelar estimações com baixos erros. Com isto, conclui-se que é plausível aplicar o critério dos mínimos quadrados a uma regressão linear, nas estimações das realizações.

Posto isto, e visto já se ter desenvolvido um programa que testa as estimações obtidas por todos os modelos possíveis de criar com duas, três e quatro variáveis, decidiu-se adaptá-lo à estimação das realizações. De notar que o banco de variáveis disponíveis para criar os ME se encontra acrescido dos indicadores relativos à construção civil, nomeadamente o IC, NNF e NNC.

Tal como no método para a estimação do IC, e visto que o programa devolveu estimações com um erro relativamente baixo, decidiu-se não aplicar outro tipo de regressão para além da linear, pelos motivos já explicados na secção anterior.

Os resultados das estimações destas variáveis obtidas pelo programa, bem como os erros associados a estas, encontram-se apresentados no Capítulo 3.

Como referido anteriormente, na metodologia para a estimação do IC, devido ao elevado número de hipóteses possíveis para obter as estimações, torna-se inexequível analisar individualmente todas as possibilidades.

2.3.4 - Investimentos Obrigatórios e Comparticipações

As comparticipações, quando se exclui o impacto das alterações na legislação, podem ser descritas unicamente em função dos IO, na medida em que seguem uma evolução similar a estes, conforme explícito na subsecção 2.1.3.

Desta forma, optou-se por realizar, primeiramente, as estimações dos IO. Para a sua estimação seguiu-se um procedimento semelhante ao aplicado às variáveis descritas anteriormente.

Com isto, começa-se por analisar a relação linear entre os IO e as variáveis disponíveis, dada pelo coeficiente de correlação de Pearson ao quadrado, ou seja, o equivalente a R^2 fornecido pela linha de tendência linear dos gráficos de dispersão. Abaixo, na Figura 2.45, pode ser visualizada, através de um gráfico de barras, a relação dos IO com as variáveis disponíveis à sua estimação. De notar que o banco de variáveis disponíveis foi acrescido das realizações, completando agora um conjunto de doze variáveis.

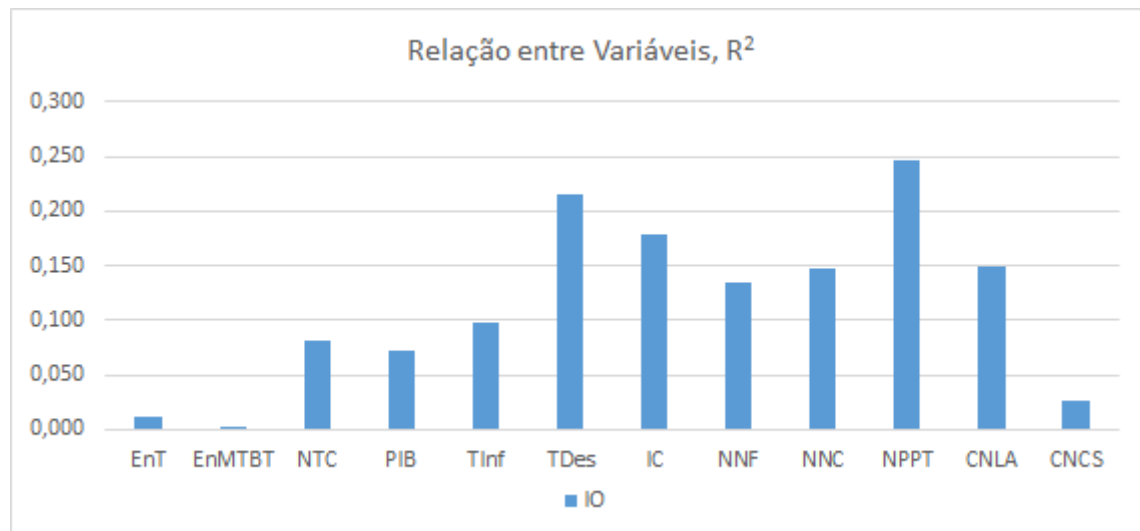


Figura 2.45 - Relação, dada por R^2 , entre as variáveis disponíveis e os IO.

Observamos que a relação linear entre as variáveis é fraca. No entanto, análogo ao sucedido para as realizações em linhas aéreas, CNLA, o agrupamento de variáveis ou mesmo as diferentes abordagens aos dados, poderão devolver estimações aceitáveis, isto é, com erros aceitáveis.

Posto isto, decidiu-se adaptar o programa para a estimação dos IO, o qual devolveu modelos com estimações bastante aceitáveis e coerentes entre si, possíveis de serem visualizadas no Capítulo 3.

Como referido, as comparticipações podem ser descritas em função dos IO, dado que estas serão sobretudo ditadas pelas diretivas regulamentares (RRC). No entanto, tendo ocorrido alterações no RRC, assumiu-se que as transições, entre os diferentes períodos sob diferente legislação, poderiam ser modeladas por uma curva de adaptação baseada em sigmóides. Desta forma, decidiu-se aplicar os conceitos descritos na subsecção 2.1.3, de forma a obter-se uma sigmoide que descreva a ocorrência de alteração no histórico, ou seja, a alteração de 2007. Visto o histórico não incluir a alteração em 2013, decidiu-se inverter esta sigmoide e moldá-la de forma a obter-se uma curva que explique esta nova alteração.

Primeiramente, decidiu-se colocar as comparticipações em função dos IO, ao dividi-las por estes, dando origem às curvas compreendidas no eixo da direita do gráfico, presente na Figura 2.46, bem como do valor absoluto, em €, dos IO, observado no eixo da esquerda.

As sigmóides de ajuste foram obtidas pela equação (2.3), presente na subsecção 2.1.3, com início em 2007, como se observa no gráfico. O parâmetro τ foi determinado pela função *Solver* do Excel, com o objetivo de minimizar o somatório dos erros quadráticos, existentes entre os valores reais e os valores devolvidos pelas sigmóides.

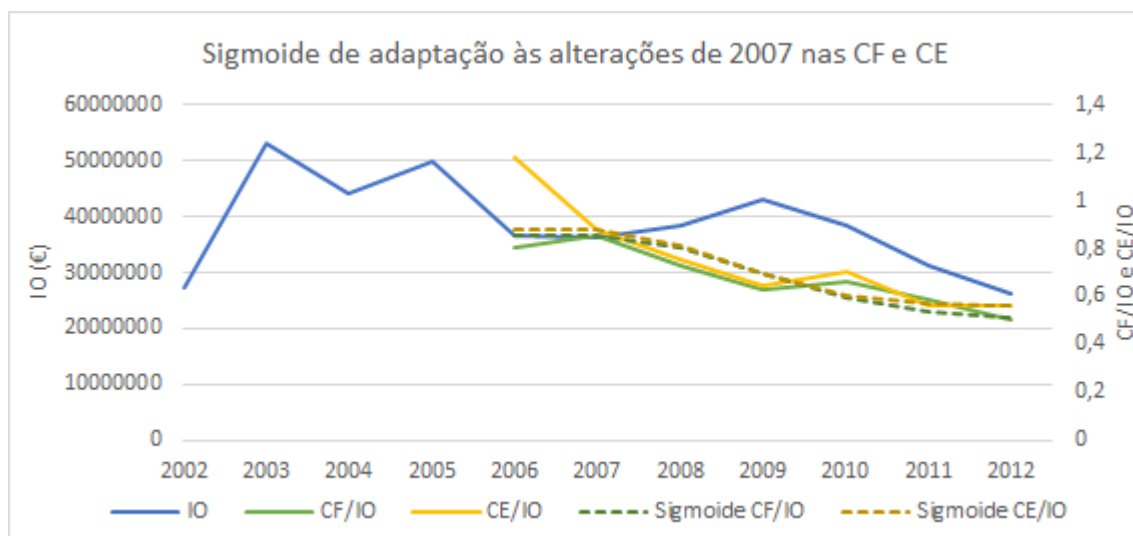


Figura 2.46 - Sigmoide decrescente de adaptação às alterações na legislação em 2007 para as CF e CE.

As sigmóides de ajuste foram obtidas pela equação (2.3), presente na subsecção 2.1.3, com início em 2007, como se observa no gráfico. O parâmetro τ foi determinado pela função *Solver* do Excel, com o objetivo de minimizar o somatório dos erros quadráticos, existentes entre os valores reais e os valores devolvidos pelas sigmóides.

A sigmoide crescente, que modeliza a alteração da legislação em 2013, foi obtida com base nos parâmetros definidos para a sigmoide decrescente. No entanto, prevê-se que as participações não voltem a atingir a mesma percentagem dos IO, que ocorria antes da alteração na legislação de 2007. Com base no parecer da EDP Distribuição, decidiu-se penalizar a amplitude máxima de variação $L_{m\acute{a}x}$, ocorrida desde 2007 até ao último ano histórico disponível, isto é, 2012, de um coeficiente k , que toma o valor de 90%.

Com a soma das sigmóides obtém-se a curva de adaptação, que, por sua vez, quando multiplicadas pelos IO, permite a obtenção das estimações das participações. Na Figura 2.47, pode visualizar-se a sigmoide crescente, bem como a curva de adaptação, para as respetivas participações:

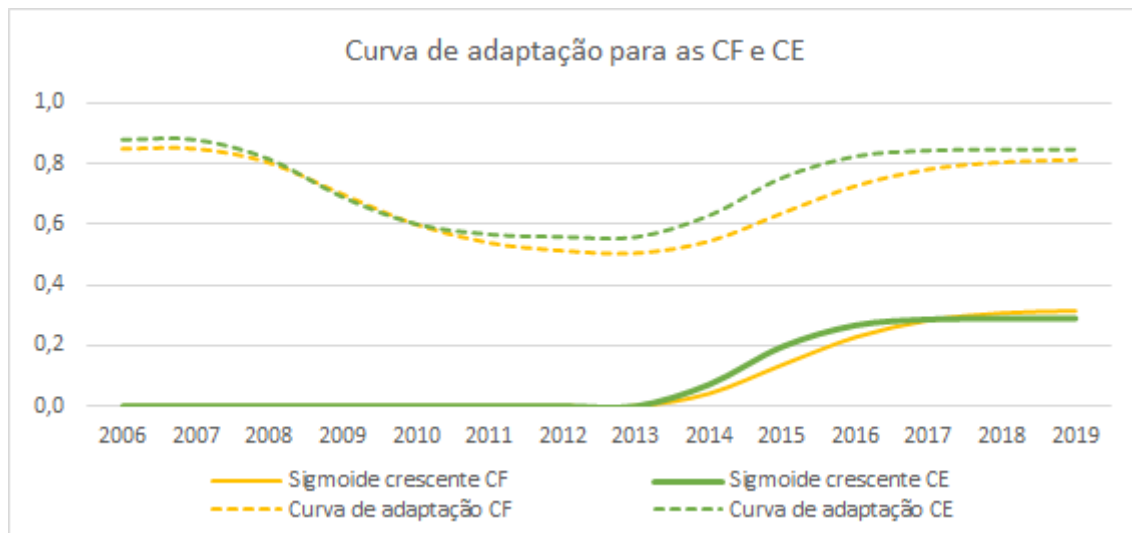


Figura 2.47 - Curva de adaptação às alterações na legislação para as CF e CE.

As estimações relativas às participações encontram-se expostas na secção 3.3.

2.4 - Programa para automatização dos Processos de estimação

A análise realizada para detetar relações entre variáveis é calculada entre pares de variáveis (por exemplo, entre IO e IC), pelo que não fornece um indicador inequívoco da importância de cada variável no ME. Isto significa que uma variável por si só pode não apresentar uma boa relação com a saída, mas quando agrupada com outras, constituindo assim um ME, pode conter uma relação admissível e gerar estimações com um erro relativamente baixo. Com isto, percebe-se que os ME não podem ser criados simplesmente com base na relação existente entre variáveis, devendo esta ser entendida simplesmente como um indicador. Além disso, estes indicadores (coeficiente de Pearson, por exemplo) apenas podem ser interpretados pela positiva, isto é, se o coeficiente é elevado, conclui-se que existe uma relação entre as duas variáveis; se o coeficiente é baixo, não se pode concluir nada.

Com um leque tão extenso de variáveis torna-se impossível testar esta relação para todos os ME possíveis de gerar. Portanto, decidiu-se desenvolver um programa em Matlab que automatize todo o processo de estimação, já referido na secção 2.3.

O fluxograma seguinte, presente na Figura 2.48, descreve a sequência dos processos desenvolvidos pelo programa. Posteriormente, encontram-se descritos os processos presentes em cada bloco do fluxograma.

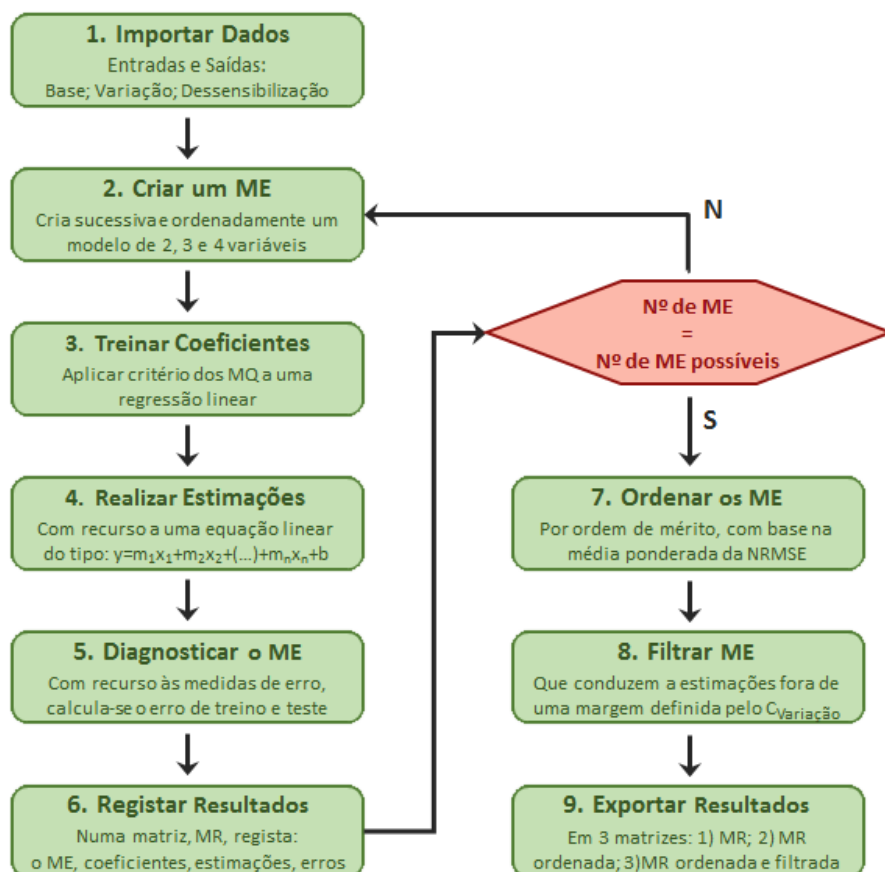


Figura 2.48 - Fluxograma de processos compreendidos no programa desenvolvido em Matlab.

1. Importar Dados

Um ficheiro Excel, utilizado como suporte e base de dados do Programa, contém os valores das três abordagens aplicadas aos dados das variáveis, ou seja, os dados de Base, os relativos à variação e os relativos à dessensibilização.

Com recurso à função do Matlab “xlsread”, importam-se estes valores para três matrizes, uma que corresponde à abordagem aos dados, outra às variáveis disponíveis para constituírem os ME e, por fim, outra para realizar as estimações. Da mesma forma, são importados os vetores referentes às saídas, ou seja, para cada uma das três abordagens.

De referir que o ciclo só termina após o 6º bloco, altura em que, se ainda não estiverem preenchidos todos os ME possíveis, é gerado um novo ME.

Nos modelos relativos às variações, os coeficientes são treinados com base na variação das saídas, pelo que se perde um dado de treino, referente ao primeiro ano de dados disponível, na medida em que esta variação é calculada com base num dado relativo ao ano anterior.

Saliente-se ainda que os vetores referentes às saídas usadas no treino de coeficientes, com dados dessensibilizados, não se encontram dessensibilizadas mas sim triplicadas, ou seja, cada dado é repetido mais duas vezes. Na Tabela 2.6, ilustra-se um exemplo do efeito descrito, no qual se pode observar como duas variáveis de entrada dessensibilizadas correspondem à saída.

Tabela 2.6 - Ilustração exemplo da correspondência de entradas dessensibilizadas com as saídas.

Ano	Entradas		Saída
	PIB (M€)	TDes (%)	IC
2002	140566,80	5,00%	119,958
	145315,06	5,88%	119,958
	135818,54	4,12%	119,958
2003	143471,70	6,30%	109,667
	148318,09	7,41%	109,667
	138625,31	5,19%	109,667
2004	149312,50	6,70%	104,900
	154356,18	7,89%	104,900
	144268,82	5,51%	104,900
2005	154268,70	7,60%	100,000
	159479,80	8,94%	100,000
	149057,60	6,26%	100,000

2. Criar um ME

Neste bloco, começa o processo iterativo do programa que consiste em criar um ME com recurso à matriz de entrada. Importa para uma matriz, referente ao ME, ordenadamente, os vetores respetivos aos dados das variáveis presentes na matriz de entrada.

De uma forma cíclica, este grava as posições das colunas importadas da matriz de entrada, no qual incrementa uma posição à posição da última coluna importada. Por sua vez, quando esta coluna atinge a posição máxima, é incrementada uma posição à coluna anterior a esta e assim sucessivamente. Quando completadas todas as hipóteses de ME possíveis de gerar, termina o ciclo e passa-se para o 7º bloco.

A título de exemplo, na Tabela 2.7, é apresentado um excerto da matriz resultados (MR), devolvida pelo programa no 6º bloco, onde se podem observar as posições das colunas na matriz de entradas, a serem importadas para o ME.

Na primeira coluna da tabela, é indicado o tipo de ME utilizado através de um código, representado por um número, que indica o número de variáveis e o tipo de abordagem aos dados. Neste exemplo, estamos perante um ME de três variáveis com dados de base. Este número indica, por ordem, o número de variáveis contidas no ME e a abordagem utilizada, ou seja, os três primeiros, valores base, variação e dessensibilização, para duas variáveis e assim sucessivamente. Note-se que é possível consultar, na Tabela 2.8, a legenda de cada código ME para cada tipo de ME criado. Saliente-se ainda que a segunda coluna indica o número do ME, isto é, a sua identidade. Por fim, é indicado a que posição da matriz de entrada se refere a coluna i, j, k e w da matriz referente ao ME criado.

Tabela 2.7 - Exemplo da posição na matriz de entrada, das colunas importadas para o ME.

Código ME	Nº de ME	Posição da coluna na matriz de entrada			
		i	j	k	w
4	46	1	2	3	
4	47	1	2	4	
...
4	49	1	2	6	
4	50	1	3	4	
...
4	55	1	5	6	
4	56	2	3	4	

Tabela 2.8 - Legenda do código referente ao tipo de ME criado.

Código ME	2 variáveis
1	Base
2	Variação
3	Dessensibilização
3 variáveis	
4	Base
5	Variação
6	Dessensibilização
4 variáveis	
7	Base
8	Variação
9	Dessensibilização

A combinação destas hipóteses dá origem, por exemplo no caso da estimação dos IO, a 2343 modelos. Como será óbvio, não é prático analisar “manualmente” todos estes casos. A automatização do processo torna-se assim uma técnica indispensável na identificação dos modelos mais adequados.

3. Treinar Coeficientes

O treino de coeficientes é feito através do critério dos mínimos quadrados [21], aplicado a uma regressão linear. Estes são obtidos com base na operação de matrizes, dada pela seguinte equação:

$$[\theta] = ([X]^T \cdot [X])^{-1} \cdot [X]^T \cdot [Z] , \quad (2.11)$$

onde:

- $[X]$, é a matriz que contém os dados de treino das variáveis do ME;
- $[Z]$, o vetor que contém os dados de treino da variável de saída;
- $[\theta]$, vetor referente aos coeficientes, com tantos argumentos quanto o número de entradas.

De referir que o coeficiente independente designado por, θ_0 , é obtido pela variável unitária incluída na matriz de entrada.

4. Realizar Estimações

Depois de obter os coeficientes realizam-se as estimações com recurso a uma equação linear, de seguida apresentada:

$$V_{Estimado} = \theta_0 + \sum_{i=1}^n \theta_i \cdot X_i \quad (2.12)$$

onde:

- $V_{Estimado}$, é o valor estimado em cada ano;
- θ_0 , é o coeficiente independente determinado pela variável unitário;
- $\sum_{i=1}^n \theta_i \cdot X_i$, somatório da multiplicação dos coeficientes pelas variáveis contidas na matriz de entrada afetas aos coeficientes.

Por exemplo, para um ME(X_1, X_2), ou seja, com duas variáveis, a equação toma a seguinte forma, $V_{Estimado} = \theta_0 + \theta_1 \cdot X_1 + \theta_2 \cdot X_2$.

As estimações são realizadas de forma iterativa para cada ano, com as entradas referentes a esse mesmo ano.

5. Diagnosticar o ME

No processo de diagnóstico do ME calculam-se as medidas de erro MAPE e NRMSE, para os anos de treino e teste, bem como a média ponderada para cada. Estes valores são guardados num vetor que, posteriormente, como descrito no próximo bloco, será anexado à matriz de resultados.

6. Registar Resultados

Neste ponto junta-se toda a informação relativa ao ME criado, isto é, o código, o número, as variáveis do ME, as estimações e, por fim, o seu diagnóstico, num vetor. A cada iteração o programa incrementa este vetor a uma nova linha da Matriz Resultados (MR). Na Tabela 2.9, ilustra-se, como exemplo, a MR para a estimação da NPPT, de forma a perceber como esta se apresenta.

Tabela 2.9 - Ilustração exemplo da MR.

Código ME	Nº do ME	Entradas				Coeficientes					Estimações			MAPE			Média			NRMSE			Média		
		i	j	k	w	θ_0	θ_1	θ_2	θ_3	θ_4	2002	...	2019	Treino	Teste	Ponderada	Treino	Teste	Ponderada	Treino	Teste	Ponderada	Treino	Teste	Ponderada
1	1	1	2			1271717	17	-30			743940	...	505793	14,6%	48,8%	35,2%	16,5%	48,8%	35,9%						
1	2	1	3			7433040	80	-2			720908	...	89035	9,7%	3,0%	5,7%	12,6%	3,0%	6,8%						
...
9	737	5	7	8	9	901413	1663362	-19	78	-18253	691538	...	597656	10,0%	59,9%	40,0%	13,1%	59,9%	41,2%						
9	738	6	7	8	9	1014947	-2108373	-10	49	-12850	729867	...	454802	9,8%	21,4%	16,8%	12,3%	21,4%	17,7%						

7. Ordenar os ME

Este bloco dá origem a uma nova matriz, que não passa da MR ordenada. Esta é ordenada por ordem de mérito e de forma crescente, em função da coluna referente à média ponderada da NRMSE.

8. Filtrar ME

Os ME que apresentam menores erros, por vezes, devolvem estimações que não se enquadram nos padrões das variáveis. De forma a eliminar estes modelos automaticamente, sem a necessidade de analisar as curvas das estimações devolvidas, decidiu-se aplicar um filtro.

O objetivo é a filtragem dos modelos que contêm estimações, referentes a anos de previsões, fora de uma certa margem, definida pelo coeficiente de variação. Em primeiro lugar, calcula-se o coeficiente de variação, isto é, o seu valor máximo e mínimo, para os últimos cinco anos do histórico, da variável a ser estimada. Com estes valores, calcula-se a margem mínima e máxima do filtro, recorrendo às seguintes equações:

$$M_{Máx} = V_{Máx} \cdot (1 + k \cdot CV) \quad (2.12)$$

$$M_{Mín} = V_{Mín} \cdot (1 - k \cdot CV) \quad (2.13)$$

onde:

- M , é a margem máxima ou mínima, respetivamente para as equações 2.12 e 2.13;
- V , é o valor máximo ou mínimo, referente aos últimos cinco anos do histórico da variável a ser estimada;
- CV , coeficiente de variação dos últimos cinco anos do histórico da variável a ser estimada;
- k , é uma constante que toma os valores de um ou dois.

Notar que o coeficiente de variação já foi abordado no processo de dessensibilização, exposto na secção 2.2.

Como será explicado mais adiante, a constante k surge pela necessidade de, por vezes, ser compensatório a nível de erros atribuir uma margem superior, pois as variáveis não se comportam todas da mesma forma. Ao utilizar dois valores de k , criam-se duas matrizes MR, as, para k igual a um e dois, respetivamente.

A título de exemplo, no gráfico exposto na Figura 2.49, podem ser observadas as estimações referentes aos dez melhores modelos devolvidos pela MR ordenada. Na mesma figura, encontram-se a tracejado azul (as duas curvas mais acima) os ME eliminados pelo filtro com $k = 1$, e a tracejado verde (curva mais em baixo) o ME eliminado pelo filtro com $k = 2$. De notar que os ME eliminados por um filtro com $k = 2$ são também eliminados com $k = 1$, pois a margem aumenta consoante o aumento do coeficiente.

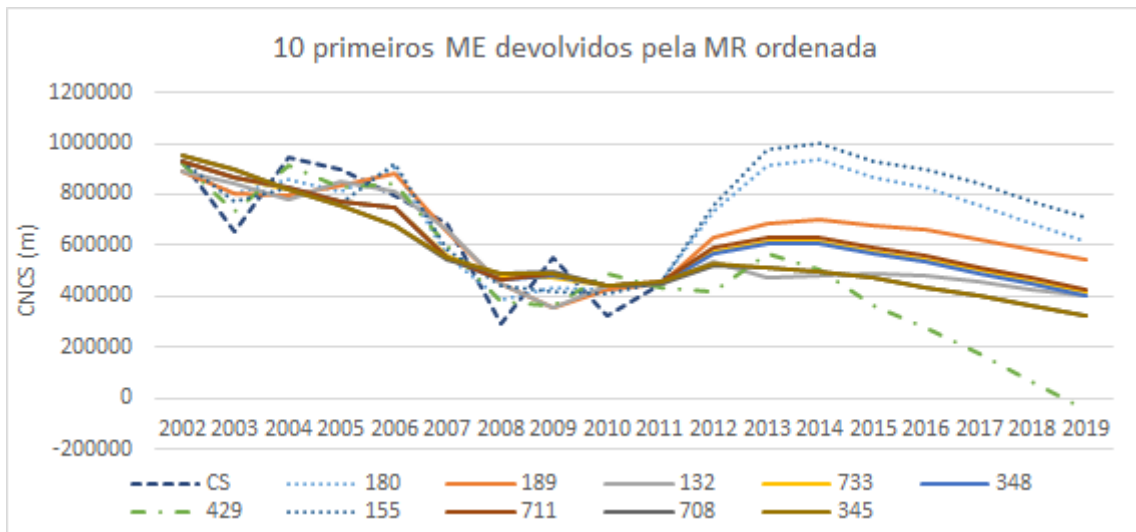


Figura 2.49 - Exemplo dos ME filtrados da MR ordenada, referente à estimação da CNCS, para uma margem com $k=1$ e $K=2$.

9. Exportar Dados

Neste último processo exportam-se as quatro matrizes criadas nos blocos seis, sete e oito, isto é, a MR, a MR ordenada e as MR ordenadas e filtradas, para k igual a um e dois. Estas foram exportadas para um ficheiro Excel, através da função Matlab “xlswrite”.

Este ficheiro Excel, previamente preparado para receber em cada folha as respetivas matrizes, realiza um gráfico com as estimações devolvidas pelos dez primeiros modelos, ou seja, no caso das matrizes ordenadas, os que possuem menores erros. Nas folhas de Excel onde se encontram as MR ordenadas e filtradas é calculada a média dos primeiros dez ME e feito o diagnóstico às estimações devolvidas por esta.

Decidiu-se realizar este último processo em ambiente Excel não só para se construir uma base de dados com um ambiente gráfico mais intuitivo, mas também para facilitar a análise gráfica dos dez melhores ME e da média resultante, devolvidos por cada MR ordenada e filtrada. Esta análise gráfica consiste em avaliar qual o valor de k que melhor se enquadra no filtro e, excecionalmente, em descartar um ou outro modelo que apresente estimações muito distantes da média, nunca se eliminando mais do que dois. De referir ainda que só são descartados os ME no caso do valor da média ponderada da NRMSE, calculada para a média das estimações, não aumentar. Por fim, as estimações são obtidas pela média dos modelos restantes, dos dez melhores modelos.

Capítulo 3

Resultados

Neste capítulo encontram-se expostas as estimações finais das variáveis necessárias de prever, relativas ao setor da construção e das realizações, de forma a melhorar o erro na estimação dos IO e das comparticipações, também aqui expostas.

São ainda indicados os ME que permitiram realizar as estimações, bem como as medidas de erro associadas a cada um destes e à estimação final. Por último, serão comentadas, individualmente, a estimação final de cada variável e, sempre que necessário, justificadas as opções que levaram a essas estimações.

Por questões de dimensão e facilidade de leitura, não se apresentam todos os resultados (tabelas, modelos, etc.). Estes encontram-se em formato eletrónico no CD que acompanha a dissertação.

3.1 - Indicadores de Construção, IC, NNF e NNC

Como se referiu na subsecção 2.3.2 a primeira variável relativa ao setor da construção a ser estimada foi o IC, através do programa desenvolvido e descrito na secção 2.4.

Estimação do IC

De seguida, na Figura 3.1, apresentam-se as estimações referentes aos dez melhores modelos presentes na MR ordenada, de forma a perceber-se qual deve ser o valor de k a introduzir no filtro, sendo esta a constante que determina a gama de valores por este aceites, ou seja, a margem mínima e máxima, como se observa nas equações (2.12) e (2.13), acima apresentadas.

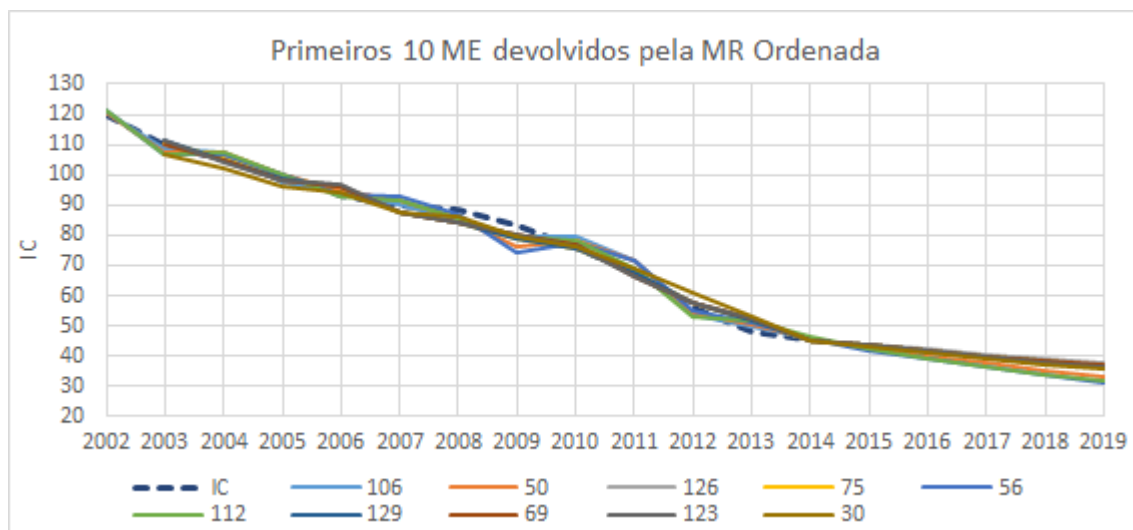


Figura 3.1 - Estimações obtidas pelos dez primeiros modelos, devolvidos pela MR ordenada, referente ao IC.

O IC, como referido na subsecção 2.1.1, onde se apresenta a variável, é caracterizado por se encontrar em défice há mais de uma década. Como a margem máxima é dada pelo máximo valor dos últimos cinco anos do histórico, entende-se que esta não é muito relevante. De referir que para uma margem calculada com $k = 2$, não é filtrado nenhum ME, ou seja, as estimações devolvidas pelos dez melhores modelos são as apresentadas na Figura 3.1, referentes à MR ordenada.

No entanto, como referido na mesma secção, as previsões disponíveis pela FEPICOP são otimistas, pelo que foi decidido limitar a margem mínima por $k = 1$. As estimações para os dez melhores ME devolvidos pela MR com um filtro definido para esta margem encontram-se expostas na Figura 3.2.

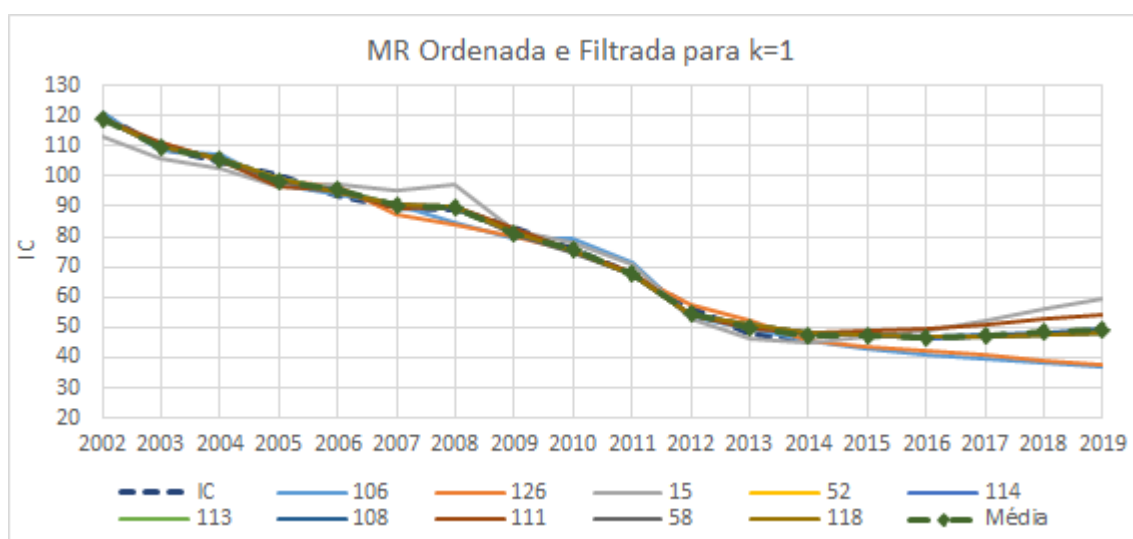


Figura 3.2 - Estimações obtidas pelos dez primeiros ME, devolvidos pela MR ordenada e filtrada para $k=1$, referente ao IC.

Como se observa, os únicos ME que resistiram ao filtro foram o 106 e o 126, o que significa que entre os presentes na Figura 3.1, estes são os que apresentam estimações mais otimistas.

As medidas de erro NRMSE relativas às estimações finais obtidas pela média dos ME devolvidos pela matriz com $k = 1$ e à média das estimações devolvidas pela matriz com um $k = 2$, podem ser observadas no seguinte gráfico de barras:

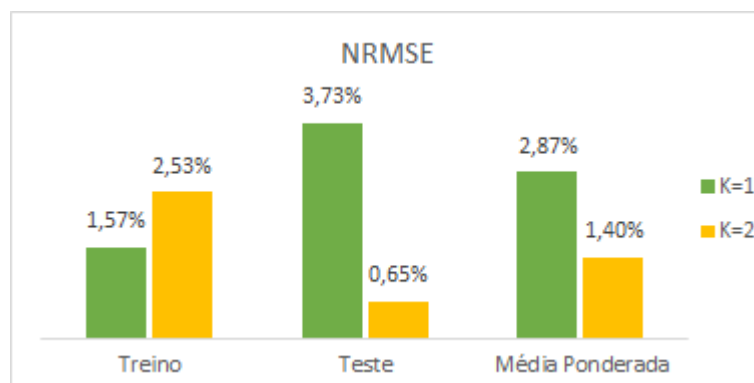


Figura 3.3 - NRMSE para a média das estimações, fornecidas pelos dez primeiros ME presentes MR ordenada e filtrada, para $k=1$ e $k=2$, na estimação do IC.

Como era de esperar, a média das estimações obtidas para $k = 2$ apresentam erros menores, principalmente no erro de teste. Isto é fácil de perceber ao observar a Figura 3.1 e Figura 3.2, sendo que na primeira as previsões mantêm a tendência do histórico, ou seja, passam muito próximo do valor real no ano de teste (2014). Contudo, como referido, prevê-se que este índice abrande a sua tendência decrescente e tome valores mais otimistas, razão pela qual se optou por uma margem definida pela constante $k = 1$.

As estimações finais para o IC foram obtidas pela média das estimações devolvidas pelos dez melhores modelos, da MR ordenada e filtrada com uma margem calculada para um $k = 1$, ou seja, a série que se designa como média presente na Figura 3.2. Esta pode ser também observada, isoladamente, na Figura 2.11.

Na Tabela 3.1 seguinte encontram-se expostos os modelos que deram origem à estimação final do IC, ou seja, dos ME presentes na Figura 3.2.

Ao analisar a média ponderada da NRMSE dos modelos presentes na tabela, em conjunto com a das estimações finais, presente na Figura 3.3, é perceptível que o erro baixa ligeiramente. Aliando este fator à consistência adquirida por juntar mais do que um modelo para a realização da estimação final, percebe-se a vantagem da aplicação deste conceito. Observa-se ainda que os ME usam predominantemente os dados de base.

Tabela 3.1 - ME usados para obtenção da estimação final do IC.

Abordagem aos dados	ME	Variáveis				NRMSE		Média Ponderada
						Treino	Teste	
Base	106	EnMTBT	EnT	NTC	PIB	2,88%	0,20%	1,27%
Variação	126	EnMTBT	EnT	TInf	TDes	3,06%	0,79%	1,70%
Base	15	TInf	TDes			4,97%	1,62%	2,96%
Base	52	EnMTBT	NTC	TDes		1,58%	4,96%	3,61%
Base	114	EnMTBT	NTC	TInf	TDes	1,58%	5,03%	3,65%
Base	113	EnMTBT	NTC	PIB	TDes	1,58%	5,04%	3,65%
Base	108	EnMTBT	EnT	NTC	TDes	1,51%	5,16%	3,70%
Base	111	EnMTBT	EnT	TInf	TDes	1,73%	5,08%	3,74%
Base	58	EnT	NTC	TDes		1,63%	5,32%	3,85%
Base	118	EnT	NTC	TInf	TDes	1,63%	5,42%	3,90%

Na Figura 3.4 e Figura 3.5, são apresentadas as estimações obtidas pelas diversas regressões para o NNF e o NNC, respetivamente. De referir que as estimações descartadas *a priori*, por apresentarem valores negativos, se encontram a tracejado.

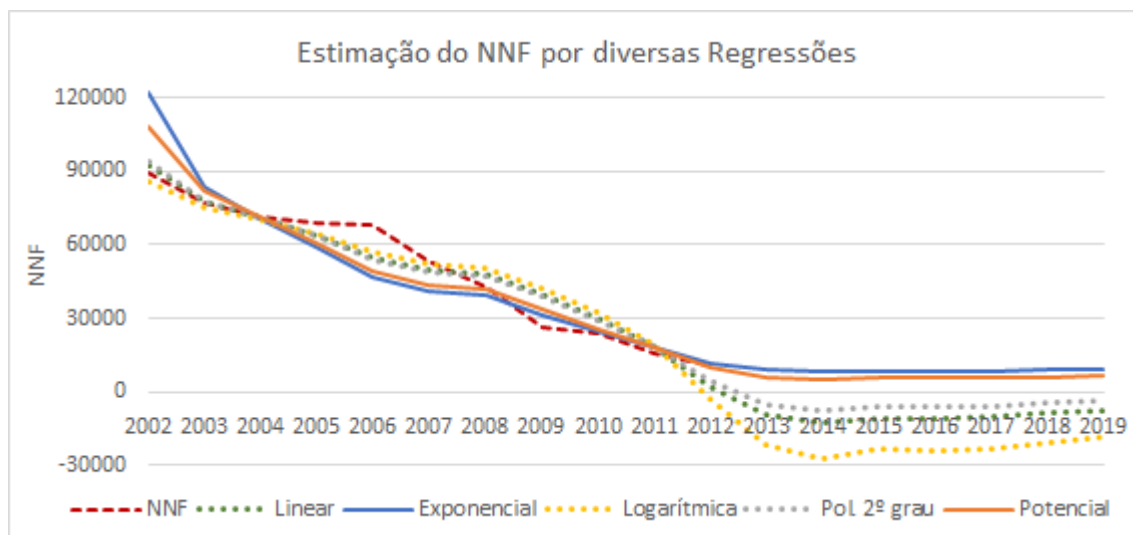


Figura 3.4 - Estimações com recurso às diversas regressões para o NNF.

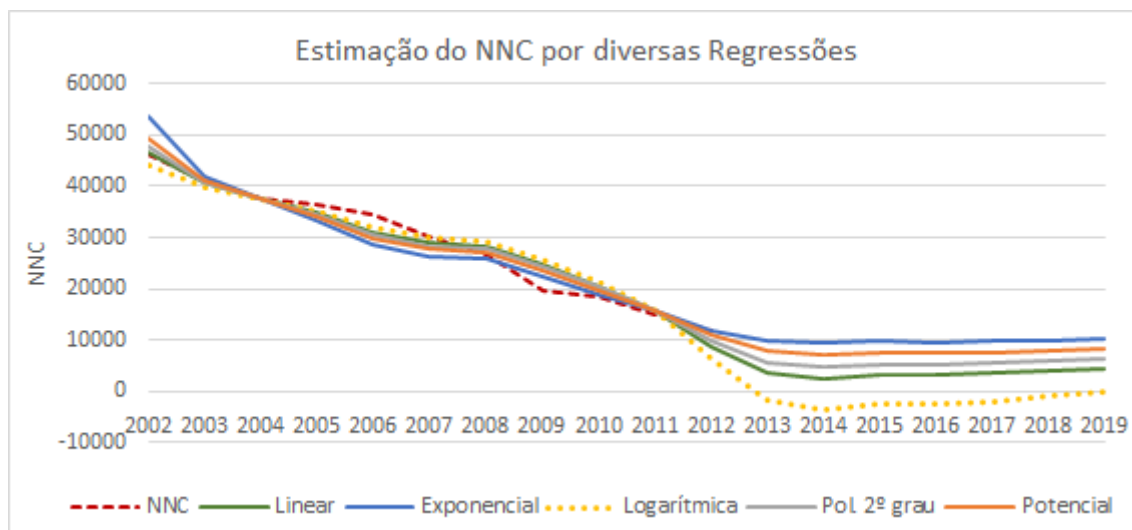


Figura 3.5 - Estimações com recurso às diversas regressões para o NNC.

Como se observa, o NNC permite ser estimado recorrendo a um maior número de regressões. Isto deve-se ao facto de estas apresentarem uma tendência decrescente menos acentuada da que ocorre para o NNF.

Nos gráficos presentes nas Figura 3.6 e Figura 3.7 são apresentados os valores da NRMSE para as respetivas regressões aceitáveis, bem como para a sua média, de forma a eleger quais as estimações finais, como descrito na subsecção 2.3.2, respetivamente para o NNF e o NNC.

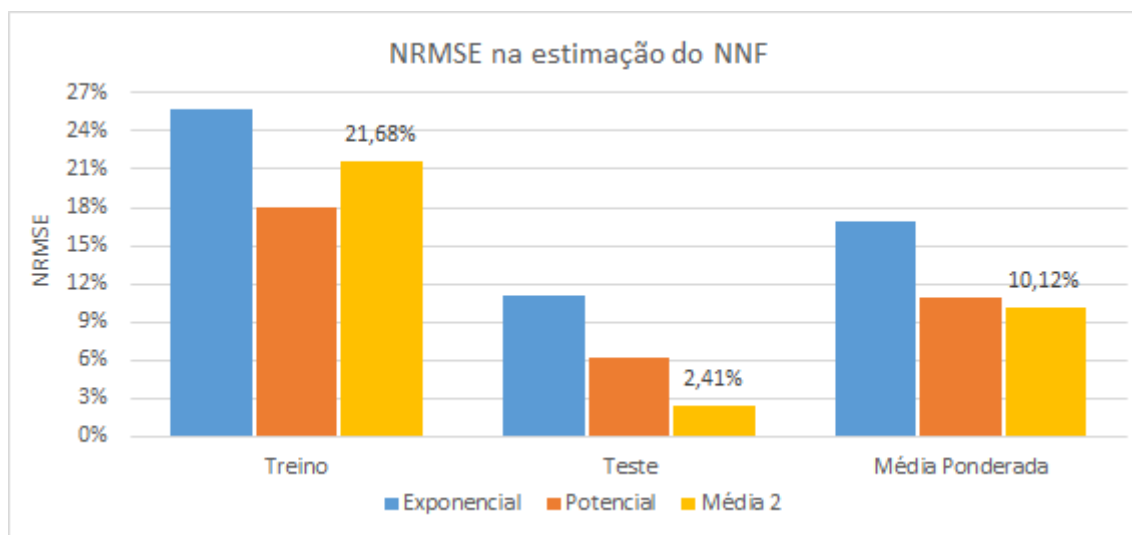


Figura 3.6 - NRMSE obtida nas estimações do NNF, pelas diversas regressões aceitáveis.

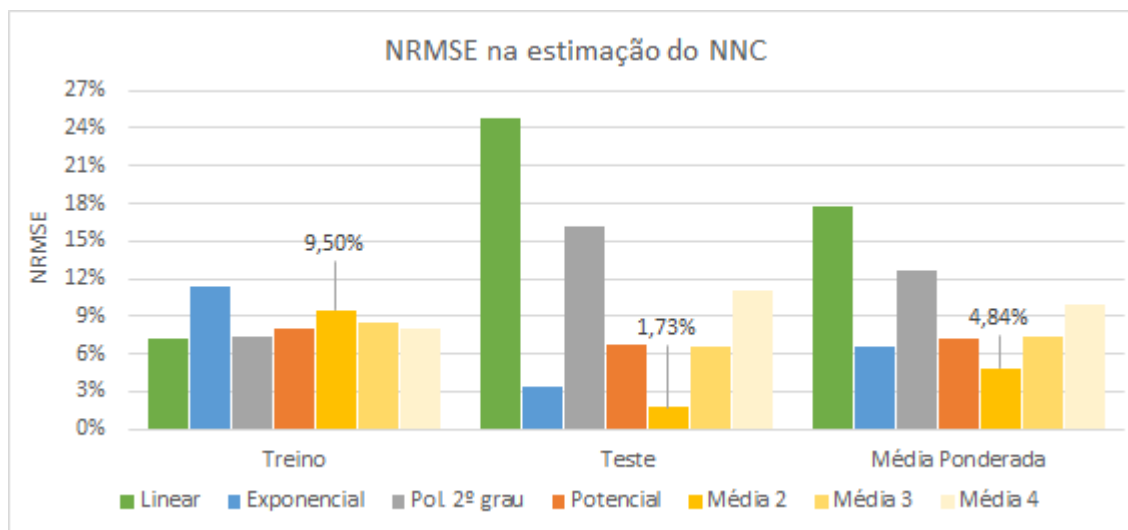


Figura 3.7 - NRMSE obtida nas estimações do NNC, pelas diversas regressões aceitáveis.

Como se observa, a hipótese que apresenta menores erros, tanto para o NNF como para o NNC, é referente à média das estimações realizadas pelas duas regressões que apresentam menor erro. Estas, coincidentemente, foram as mesmas para os dois casos, ou seja, a exponencial e a potencial, sendo que para o NNF eram as únicas disponíveis. Constata-se ainda que o erro referente ao NNF é sensivelmente o dobro do obtido na estimação do NNC.

As estimações finais para o NNF e NNC podem ser observadas nas Figura 2.12 e Figura 2.13, respetivamente. Por sua vez, os valores absolutos da NRMSE obtidos nas estimações finais do NNF e NNC podem ser observados nas Figura 3.6 e Figura 3.7, respetivamente. De referir que, como era de esperar, estes indicadores apresentam uma tendência semelhante ao IC.

3.2 - Realizações EDP Distribuição: NPPT, CNLA e CNCS

Como referido na subsecção 2.3.3, procedeu-se às estimações destas variáveis recorrendo ao programa desenvolvido e descrito na secção 2.4.

Primeiramente, serão analisadas as estimações devolvidas pelos dez primeiros modelos presentes na matriz MR ordenada, ou seja, sem a aplicação do filtro. Com isto, é possível observar-se a tendência das melhores estimações devolvidas pelo programa, o que torna mais fácil decidir qual a margem a aplicar no filtro, sem descartar eventuais modelos que poderiam indicar uma tendência descartada pelo filtro. De referir ainda que só em casos justificados se opta por um filtro com uma margem inferior, ou seja, para $k = 1$, pois ao diminuir a margem podem eliminar-se ME que apresentam uma média ponderada da NRMSE, de treino e teste, inferior.

Posteriormente, serão apresentadas as estimações referentes aos dez melhores modelos devolvidos pela MR ordenada e filtrada, para os diferentes valores de k . Por fim, serão apresentados os ME que deram origem às estimações finais, obtidas pela sua média.

Decidiu-se apresentar as variáveis isoladamente, de forma a tornar mais perceptível a sua leitura, pela ordem com que foram apresentadas, isto é, NPPT, CNLA e CNCS.

Na análise destas variáveis é importante relembrar, como referido na subsecção 2.1.1 aquando da apresentação destas, que os picos existentes no seu histórico advêm da possibilidade de, por vezes, parte das realizações serem contabilizadas em anos posteriores.

Estimação da NPPT

No gráfico seguinte são apresentadas as estimações devolvidas pelos primeiros modelos presentes na MR ordenada.

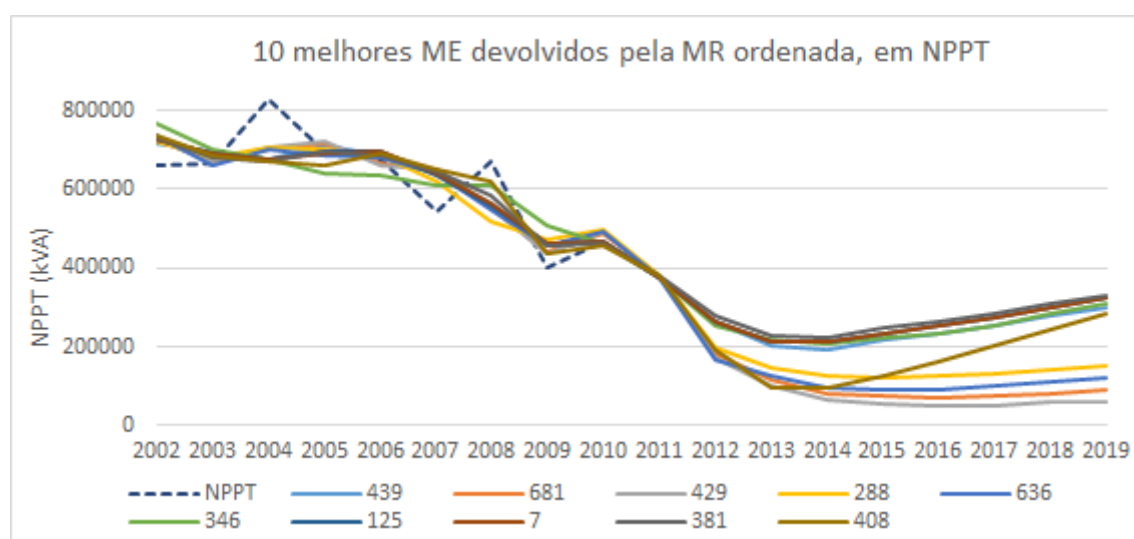


Figura 3.8 - Estimações obtidas pelos dez primeiros modelos, devolvidos pela MR ordenada, referente à NPPT.

Como referido, as realizações advêm da necessidade de reforçar as redes. No entanto, observa-se que as estimações apresentam uma tendência pessimista e contraditória ao que se verifica nos restantes índices, na medida em que se prevê uma melhoria da economia, do setor industrial e da construção civil. Com isto, percebe-se a importância de aplicar um filtro.

Nas seguintes figuras apresentam-se as MR ordenadas e filtradas para os diferentes valores de k , isto é, 2 e 1, respetivamente.

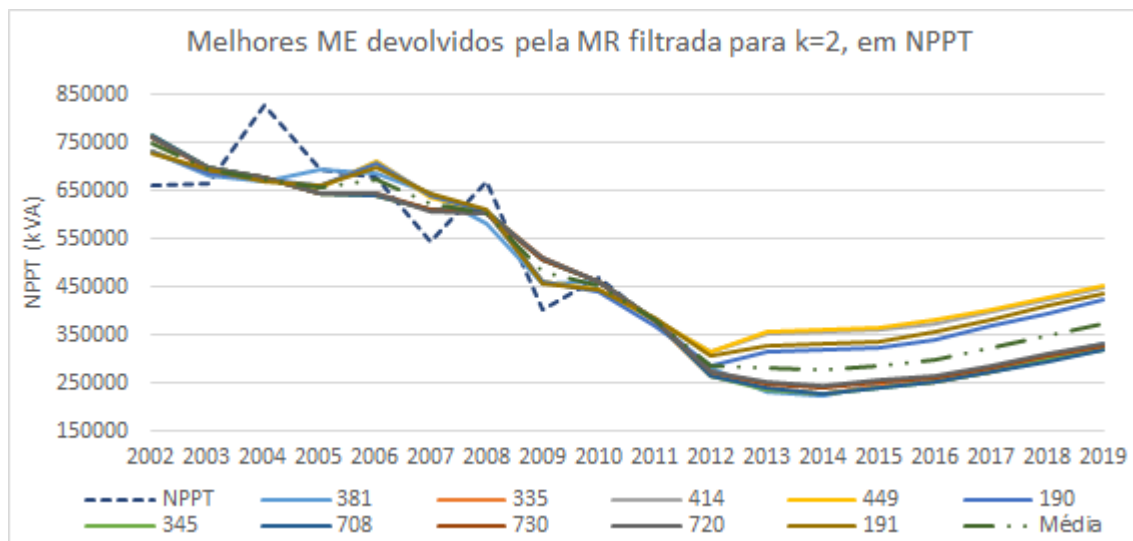


Figura 3.9 - Estimações obtidas pelos dez primeiros ME, devolvidos pela MR ordenada e filtrada para $k=2$, referente à NPPT.

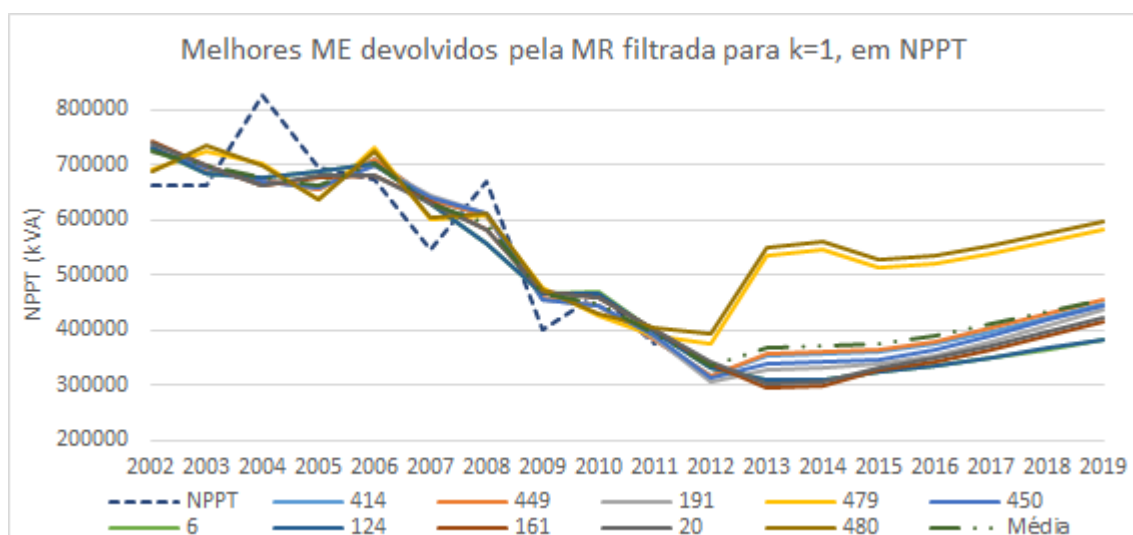


Figura 3.10 - Estimações obtidas pelos dez primeiros ME, devolvidos pela MR ordenada e filtrada para $k=1$, referente à NPPT.

Os valores da NRMSE obtidos para a média das estimações dos dez melhores modelos, devolvidos pela MR ordenada e filtrada, para os diferentes valores de k , podem ser observados na figura seguinte:

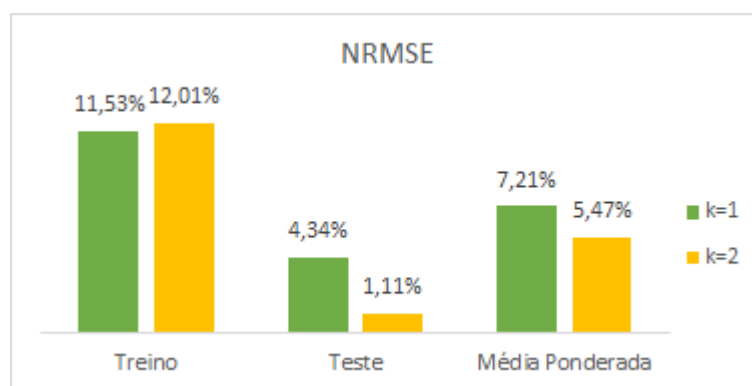


Figura 3.11 - NRMSE para a média das estimações, fornecidas pelos dez primeiros ME presentes MR ordenada e filtrada, para $k=1$ e $k=2$, na estimação da NPPT.

Como se pode observar, as estimações obtidas com $k = 1$ apresentam uma NRMSE superior e com maior discrepância nas estimações devolvidas por estes modelos. Posto isto, junto dos conceitos afetos a esta variável, anteriormente referidos, e depois de analisar as curvas das estimações presentes nas Figura 3.9 e Figura 3.10, definiu-se as estimações finais como a média das estimações dos modelos devolvidos pela MR ordenada e filtrada, com uma margem definida por $k = 2$.

As estimações finais podem ser observadas pela série que se designa como média presente na Figura 3.9 ou, isoladamente, na Figura 2.14 (capítulo onde se apresentam as variáveis).

Na tabela seguinte encontram-se expostos os modelos que deram origem às estimações finais, ou seja, dos ME presentes na Figura 3.9.

Tabela 3.2 - ME usados para obtenção da estimação final da NPPT.

Abordagem aos dados	ME	Variáveis				NRMSE		Média Ponderada
						Treino	Teste	
Base	381	EnMTBT	EnT	NNC	IC	12,08%	0,98%	5,42%
Dessensibilização	335	NTC	TDes	NNF		12,82%	0,60%	5,49%
Base	414	EnMTBT	TDes	NNF	IC	11,68%	1,50%	5,57%
Base	449	ENT	TDes	NNF	IC	11,67%	1,57%	5,61%
Base	190	TDes	NNF	IC		11,68%	1,66%	5,67%
Dessensibilização	345	PIB	TDes	NNF		12,83%	0,93%	5,69%
Dessensibilização	708	NTC	PIB	TDes	NNF	12,84%	1,48%	6,02%
Dessensibilização	730	TInf	TDes	NNF	NNC	12,75%	1,60%	6,06%
Dessensibilização	720	NTC	TDes	NNF	NNC	12,75%	1,61%	6,06%
Base	191	TDes	NNC	IC		11,73%	2,47%	6,18%

Semelhante ao que ocorre para o IC, ao analisar a média ponderada da NRMSE dos modelos presentes na tabela, em conjunto com a das estimações finais, obtida pela média das estimações fornecidas pelos modelos, percebe-se a vantagem da aplicação deste conceito. É de notar que os indicadores de construção, ao estarem presentes em todos os ME selecionados, se revelaram importantes na estimação desta variável. De referir ainda que mesmo não

existindo grande correlação individual, entre a NPPT e as restantes variáveis disponíveis, ao agrupá-las, foi possível obter estimações bastante aceitáveis e com erros relativamente baixos.

Estimação da CNLA

No gráfico seguinte são expostas as estimações devolvidas pelos dez primeiros ME presentes na MR ordenadas.

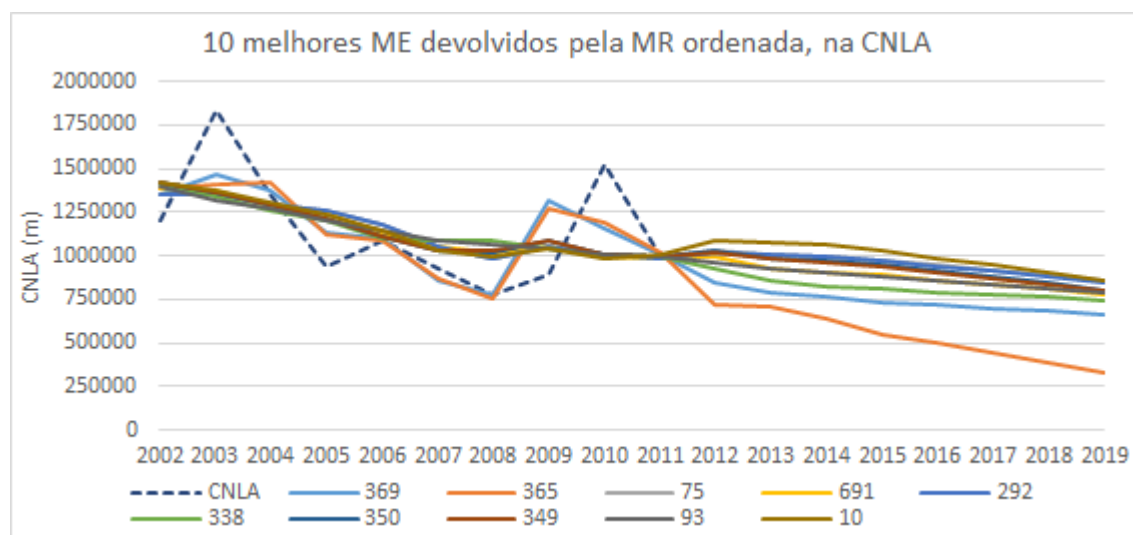


Figura 3.12 - Estimações obtidas pelos dez primeiros modelos, devolvidos pela MR ordenada, referente à CNLA.

Como se observa, os ME têm grande dificuldade em se adaptarem à curva real, pelo que produzem grandes erros de treino. Isto deve-se, mais uma vez, aos picos presumivelmente inexplicáveis pelas restantes variáveis, existentes em alguns anos, neste caso em 2003 e 2010.

Verifica-se também que os primeiros ME devolvidos com a exceção do 369 e 365 usam as suas entradas dessensibilizadas, por conseguirem desvalorizar, de certa forma, os picos existentes.

Visto que as MR ordenadas e filtradas para os diferentes valores de k devolvem, igualmente, os primeiros dez ME, só será apresentado um dos casos ($k = 2$).

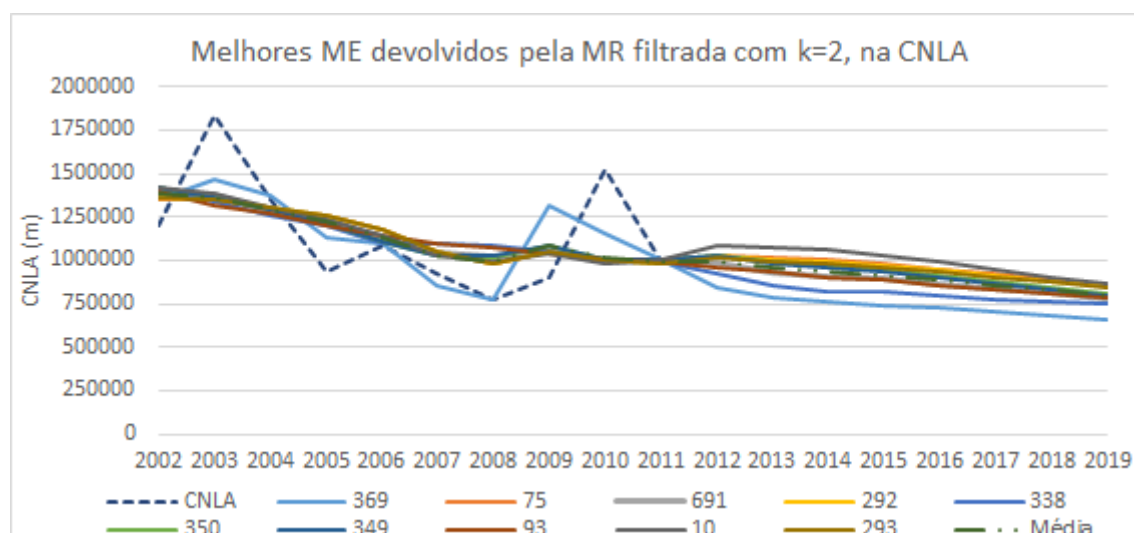


Figura 3.13 - Estimações obtidas pelos dez primeiros ME, devolvidos pela MR ordenada e filtrada para $k=2$, referente à CNLA.

Note-se que de forma a eliminar o peso do erro de teste na ordenação da MR, experimentou-se calcular a média ponderada apenas com vinte por cento do erro de teste e analisar a MR ordenada e filtrada com a mesma margem, isto é, para $k = 2$. No entanto, a média das estimações dos modelos por esta devolvidos, como se observa na figura seguinte, é idêntica à apresentada na Figura 3.13.

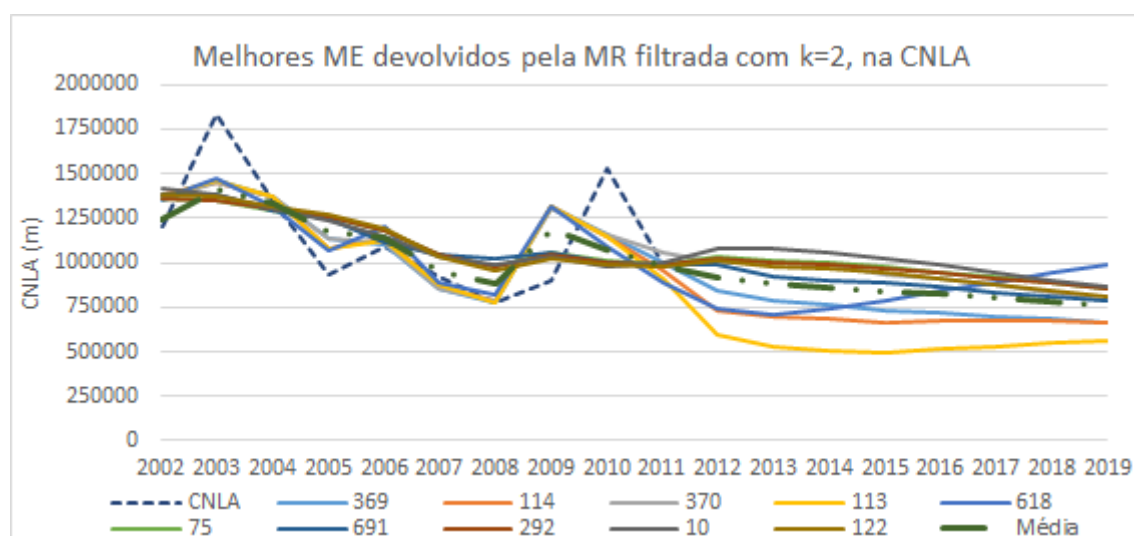


Figura 3.14 - Estimações obtidas pelos dez primeiros ME, devolvidos pela MR ordenada por uma média ponderada calculada com apenas 20% de teste, referente à CNLA.

Os valores obtidos para a NRMSE, nas diferentes abordagens referidas, são expostos na figura seguinte. De referir que o cálculo da média ponderada com vinte por cento de teste apenas foi utilizada no processo de ordenação da MR, sendo a média ponderada apresentada no gráfico calculada da forma habitual, de forma a permitir uma comparação justa.

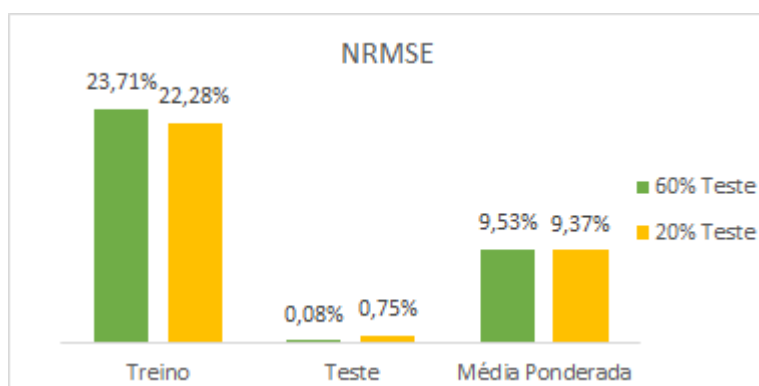


Figura 3.15 - NRMSE para a média das estimações fornecidas pelos dez melhores ME obtidos pela MR ordenada, para as diferentes formas de cálculo da média ponderada, referente às CNLA.

Efetivamente, verifica-se uma ligeira diminuição da NRMSE no treino e na média ponderada. Contudo, visto que as previsões dadas pela estimacão são idênticas nos dois casos e a diminuição do erro é mínima, optou-se por utilizar as estimacões obtidas da forma habitual.

No entanto, fica registada uma estratégia possível de trazer melhorias em estimacões com erros semelhantes a estes, ou seja, que possuam um erro de treino muito superior ao de teste.

As estimacões finais para a CNLA podem ser observadas pela série média presente na Figura 3.13 ou, isoladamente, na Figura 2.15.

Na tabela seguinte, encontram-se expostos os modelos que deram origem às estimacões finais:

Tabela 3.3 - ME usados para obtenção da estimacão final da CNLA.

Abordagem aos dados	ME	Variáveis				NRMSE		Média Ponderada
						Treino	Teste	
Base	369	EnMTBT	EnT	PIB	NNF	20,41%	0,89%	8,70%
Dessensibilizacão	75	EnMTBT	PIB			24,19%	0,20%	9,79%
Dessensibilizacão	691	EnT	PIB	NNF	NNC	24,20%	0,43%	9,94%
Dessensibilizacão	292	EnMTBT	PIB	NNF		24,24%	0,70%	10,12%
Dessensibilizacão	338	NTC	NNF	NNC		25,20%	0,11%	10,15%
Dessensibilizacão	350	PIB	NNC	IC		24,36%	0,68%	10,15%
Dessensibilizacão	349	PIB	NNF	IC		24,33%	0,80%	10,21%
Dessensibilizacão	93	NTC	IC			25,43%	0,23%	10,31%
Base	10	EnT	PIB			24,22%	1,07%	10,33%
Dessensibilizacão	293	EnMTBT	PIB	NNC		24,29%	1,25%	10,46%

Como referido, devido à dificuldade de explicar os picos existentes no histórico da CNLA e também da baixa correlacão desta com as restantes variáveis, os ME seleccionados usam, na sua maioria, entradas dessensibilizadas. Posto isto, é possível explicar os elevados erros de treino, pelo que é necessário ter em conta a inclusão desta variável na estimacão de outras variáveis. Porém, apesar de um erro superior de treino, consegue-se obter para as suas previsões um indicador médio, que se espera próximo da realidade, semelhante ao que sucede no histórico.

Estimação da CNCS

Na Figura 2.49, anteriormente apresentada, encontram-se expostas as estimações devolvidas pelos primeiros dez modelos presentes na MR ordenadas. Como se observa, os modelos apresentam uma grande disparidade nas suas estimações, pelo que se revela imprescindível a aplicação do filtro.

As figuras seguintes apresentam as MR ordenadas e filtradas para os diferentes valores de k , isto é, 2 e 1, respetivamente.

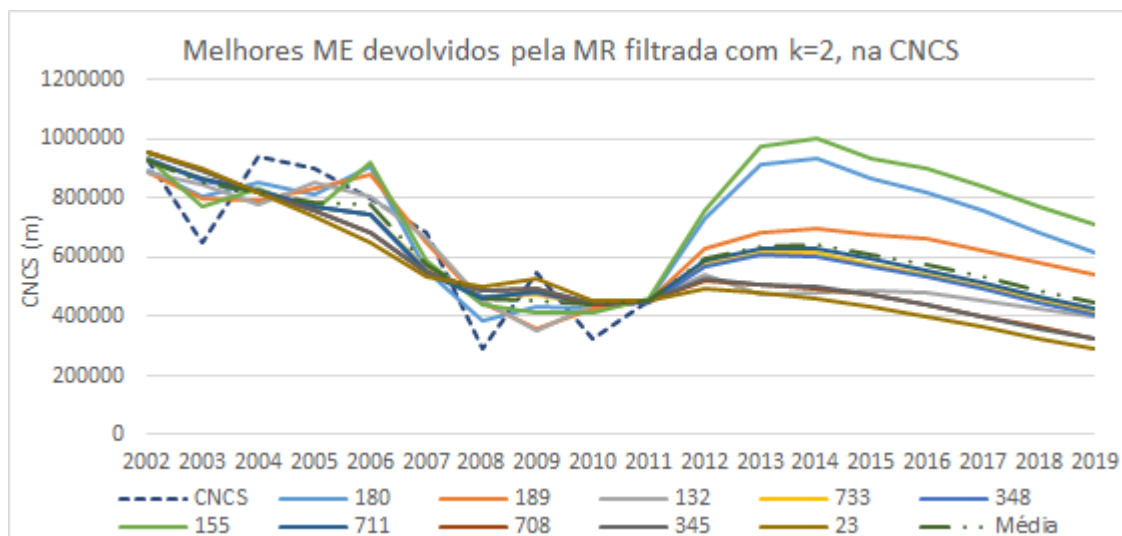


Figura 3.16 - Estimações obtidas pelos dez primeiros ME, devolvidos pela MR ordenada e filtrada para $k=2$, referente à CNCS.

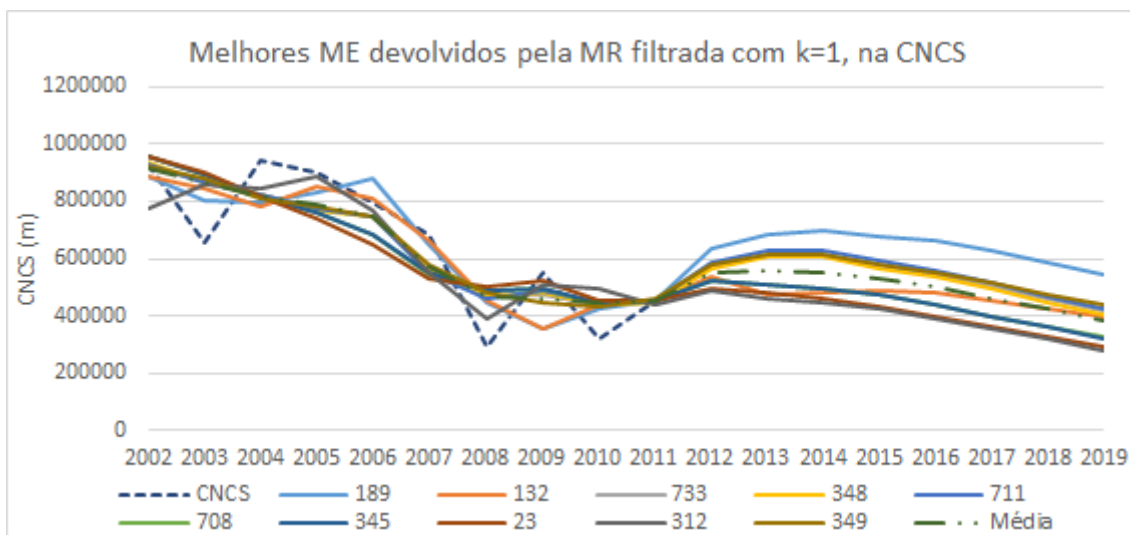


Figura 3.17 - Estimações obtidas pelos dez primeiros ME, devolvidos pela MR ordenada e filtrada para $k=1$, referente à CNCS.

Os valores da NRMSE obtidos para a média das estimações dos dez melhores modelos devolvidos pela MR ordenada e filtrada, para os diferentes valores de k , podem ser observados na figura abaixo:

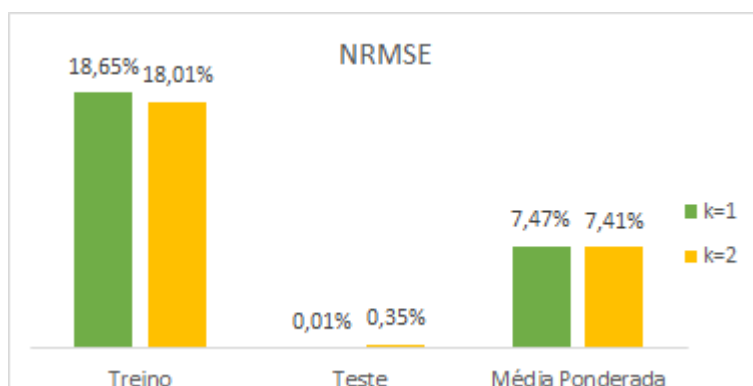


Figura 3.18 - NRMSE para a média das estimações, fornecidas pelos dez primeiros ME presentes MR ordenada e filtrada, para $k=1$ e $k=2$, na estimação do CNCS.

Com a obtenção de erros muito idênticos e visto os modelos obtidos para um $k = 1$ apresentarem estimações mais concisas entre si, ou seja, com um desvio em relação à média inferior, optou-se por utilizar os modelos devolvidos por este, presentes na Figura 3.17, para adquirir as estimações finais, através da média das suas estimações.

As estimações finais podem ser observadas pela série que se designa como média presente na Figura 3.17 ou, isoladamente, na Figura 2.16.

Na tabela seguinte, encontram-se expostos os modelos que deram origem às estimações finais:

Tabela 3.4 - ME usados para obtenção da estimação final da CNCS.

Abordagem aos dados	ME	Variáveis				NRMSE		Média Ponderada
						Treino	Teste	
Base	189	TDes	NNF	NNC		17,92%	0,01%	7,18%
Base	132	EnMTBT	TDes	NNC		18,93%	0,02%	7,58%
Dessensibilização	733	PIB	NNF	NNC	IC	18,88%	0,34%	7,76%
Dessensibilização	348	PIB	NNF	NNC		18,83%	0,42%	7,78%
Dessensibilização	711	NTC	PIB	NNF	NNC	18,84%	0,89%	8,07%
Dessensibilização	708	NTC	PIB	TDes	NNF	21,18%	0,07%	8,52%
Dessensibilização	345	PIB	TDes	NNF		21,19%	0,43%	8,74%
Base	23	PIB	TDes			22,61%	0,13%	9,12%
Dessensibilização	312	EnT	PIB	TDes		18,17%	3,09%	9,12%
Dessensibilização	349	PIB	NNF	IC		19,54%	2,19%	9,13%

De forma semelhante ao que sucede para a CNLA, os ME seleccionados usam como entradas dados dessensibilizados, o que se deve essencialmente aos mesmos fatores. Porém, esta variável apresenta erros relativamente inferiores, pelo que se considera aceitável incluí-la sem restrições na estimação de outras variáveis.

De forma análoga ao que sucede para as CE suspeita-se que, por vezes, parte das realizações só serão contabilizadas em anos posteriores. Acrescentando a este o fator “anos de eleições”, que explica em parte alguns picos do histórico dos IO, compreende-se a dificuldade existente na aproximação das estimações devolvidas pelos modelos aos valores históricos.

3.3 - Investimentos Obrigatórios e Comparticipações

Adquiridas todas as variáveis disponíveis à estimação dos IO, um dos objetivos principais do trabalho desenvolvido, procedeu-se à sua estimação através do programa desenvolvido.

Com uma estrutura semelhante à apresentação de resultados das restantes variáveis estimadas, serão expostas as estimações finais dos IO.

Posteriormente, como descrito na subsecção 2.3.4, expõem-se as estimações finais das comparticipações obtidas pelos IO.

Estimação dos IO

Semelhante à exposição de resultados das variáveis estimadas pelo programa, são abaixo apresentadas as estimações referentes aos dez melhores modelos presentes na MR ordenada.

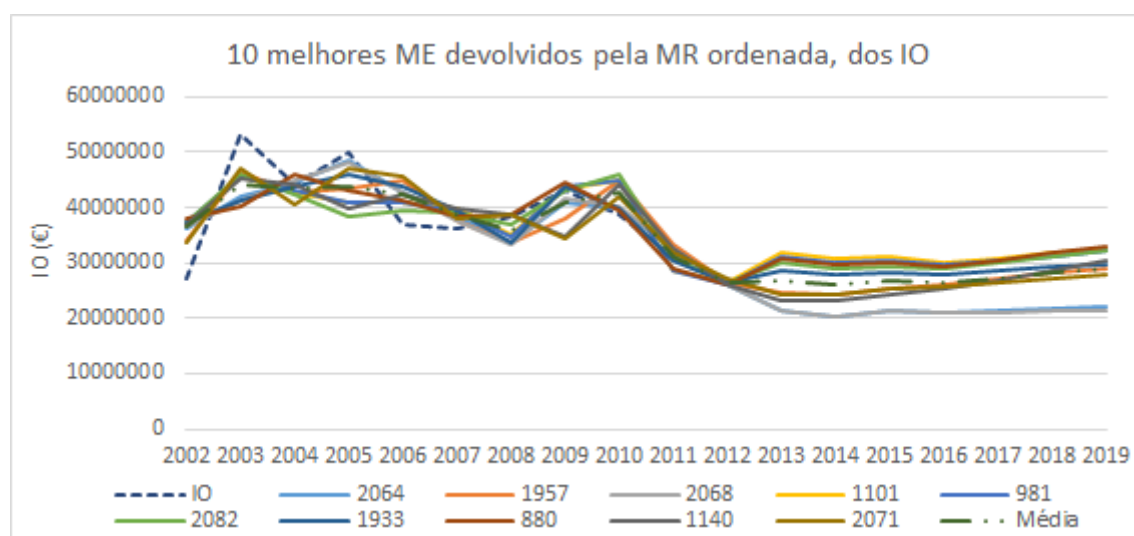


Figura 3.19 - Estimações obtidas pelos dez primeiros modelos, devolvidos pela MR ordenada, referente aos IO.

Observa-se que os ME apresentam, no ano de teste (2012), uma boa aproximação ao valor histórico, possivelmente, por serem ordenados pela média ponderada da NRMSE com um peso superior no teste.

Como os IO se encontram em queda há quatro anos, a margem máxima a aplicar no filtro dos ME não terá grande relevância. Contudo, tal como no caso do IC, é importante ter em atenção qual a margem mínima a aplicar no filtro. No entanto, como está previsto que a queda

na atividade do setor da construção estagne e apresente, inclusive, sinais de melhoria nos próximos anos, semelhante à atividade económica do país e ao aumento do consumo de energia, também aqui se torna importante limitar a margem mínima.

Para uma margem calculada com $k = 2$ não é filtrado nenhum ME da MR ordenada, pelo que se decidiu apresentar a média das estimações devolvidas pelos modelos expostos na Figura 3.19, referente a esta.

Na figura seguinte são apresentados os primeiros dez ME devolvidos pela MR ordenada, filtrada para uma margem definida por $k = 1$.

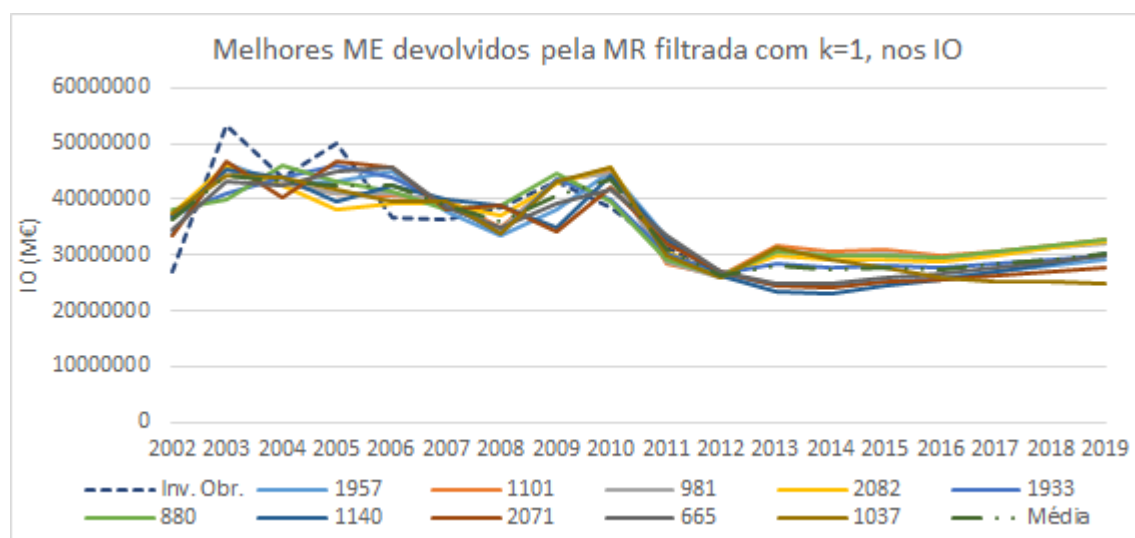


Figura 3.20 - Estimações obtidas pelos dez primeiros ME, devolvidos pela MR ordenada e filtrada para $k=1$, referente aos IO.

Como se observa, só foram retidos pelo filtro dois ME, o 2064 e o 2068, os quais apresentam previsões mais baixas, e incluídos os ME 665 e 1037.

As medidas de erro NRMSE relativas à média das estimações devolvidas pela matriz com um $k = 1$ e $k = 2$, podem ser observadas no gráfico de barras seguinte:

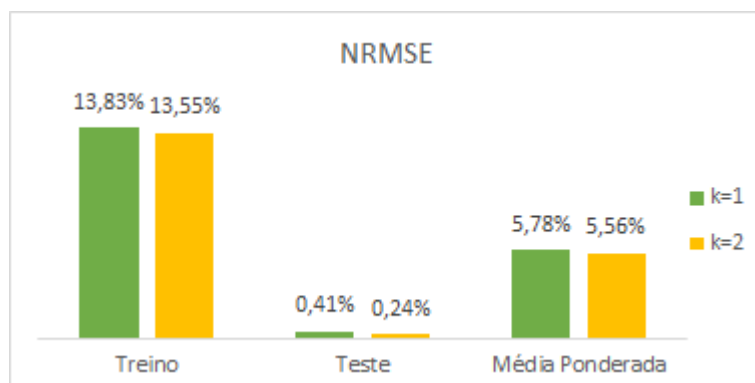


Figura 3.21 - NRMSE para a média das estimações, fornecidas pelos dez primeiros ME presentes MR ordenada e filtrada, para $k=1$ e $k=2$, na estimação dos IO.

Como a NRMSE obtida para os dois casos é muito semelhante, torna-se necessário observar com atenção a média das estimações dos modelos para os dois casos, isto é, com uma margem a aplicar no filtro calculada com $k = 1$ e $k = 2$, as quais podem ser observadas isoladamente abaixo.

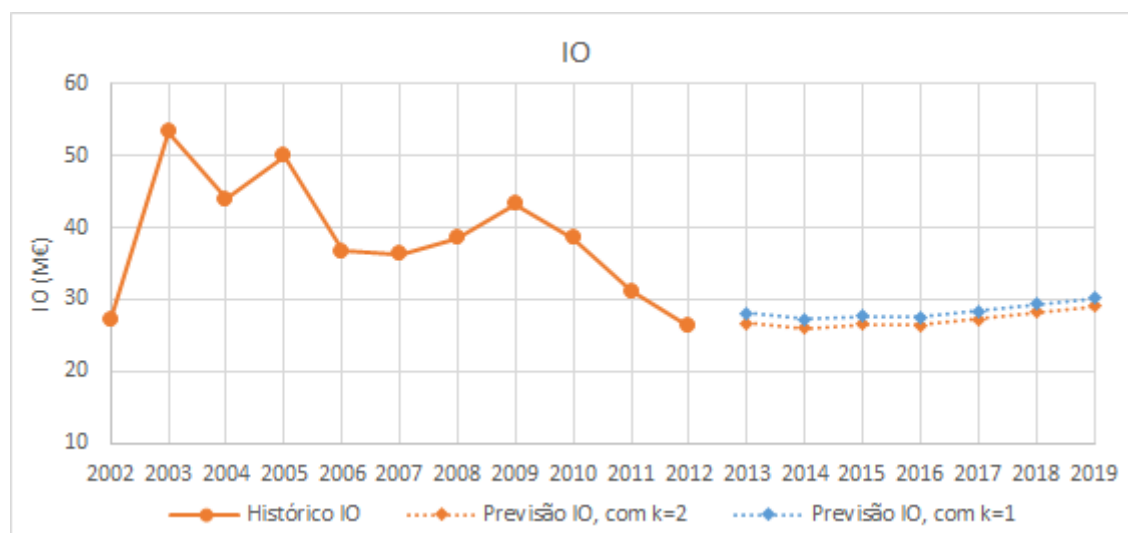


Figura 3.22 - Histórico e Previsão dos IO.

Sendo as estimações similares, optou-se como estimações finais as obtidas pela média das estimações devolvidas pelos dez melhores modelos da MR ordenada e filtrada, com uma margem calculada para um $k = 2$, as quais apresentam uma NRMSE para treino e teste inferior, apesar de ser uma diferença mínima.

Na tabela seguinte, encontram-se expostos os modelos que deram origem à estimação final dos IO, ou seja, dos ME presentes na Figura 3.19.

Tabela 3.5 - ME usados para obtenção da estimação final dos IO.

Abordagem aos dados	ME	Variáveis				NRMSE		Média Ponderada
						Treino	Teste	
Dessensibilização	2064	EnT	PIB	NNF	IC	13,47%	0,46%	5,67%
Dessensibilização	1957	EnMTBT	PIB	CNLA	CNCS	13,86%	0,71%	5,97%
Dessensibilização	2068	EnT	PIB	NNC	IC	13,88%	0,90%	6,09%
Base	1101	EnT	TInf	NNC	CNLA	14,38%	0,65%	6,14%
Base	981	EnMTBT	TInf	NNC	CNLA	14,51%	0,74%	6,25%
Dessensibilização	2082	EnT	TInf	TDes	CNLA	15,20%	0,65%	6,47%
Dessensibilização	1933	EnMTBT	PIB	TInf	IC	14,82%	0,94%	6,49%
Base	880	EnMTBT	EnT	TInf	NPPT	15,39%	0,58%	6,51%
Base	1140	EnT	IC	NPPT	CNLA	16,15%	0,57%	6,80%
Dessensibilização	2071	EnT	PIB	NNC	CNCS	13,21%	2,70%	6,90%

Coincidência ou não, verifica-se que os anos com picos, isto é, 2005 e 2009, são coincidentes com anos de eleições autárquicas, pelo que se crê que numa nova abordagem

estes devam ser incluídos aquando da realização das estimações para os IO. No entanto, com um histórico tão reduzido não é possível retirar conclusões definitivas sobre estas relações.

De notar ainda que mesmo existindo uma baixa correlação dos IO com as restantes variáveis, como observado na Figura 2.45, estas, quando relacionadas em conjunto (a saída passa a depender do produto de duas ou mais entradas), conseguem ajustar-se à curva definida pelo histórico, comprovando a possibilidade descrita na subsecção 2.3.1.

Refira-se ainda que foram também testados modelos sem inclusão da CNLA nas variáveis de entrada, dado estas apresentarem um erro considerável. No entanto, as estimações finais foram bastante parecidas e os erros ligeiramente superiores, pelo que se considerou desnecessário apresentar estes resultados em detalhe.

Estimação das CF e CE

Como mencionado na subsecção 2.3.4, tendo em conta as alterações na legislação referidas na subsecção 2.1.3, as estimações das participações foram realizadas unicamente através dos IO com recurso às curvas de adaptação presentes na Figura 2.47. De notar a necessidade de terem sido apresentados os resultados aquando da exposição da metodologia para uma melhor perceção.

Multiplicando as respetivas curvas de adaptação pelos IO, obtiveram-se as estimações presentes na figura abaixo, para as CF e CE.

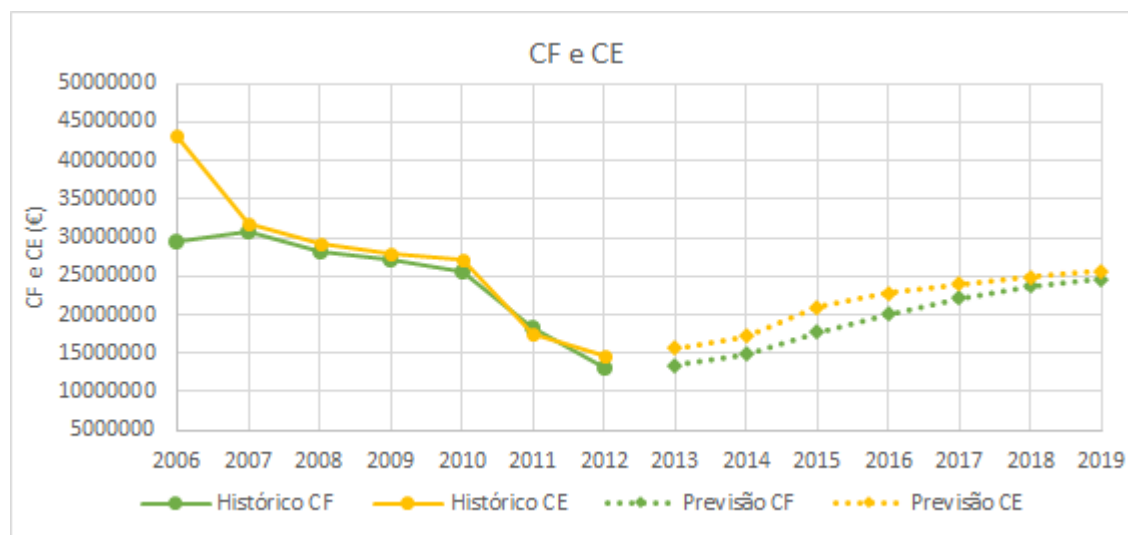


Figura 3.23 - Histórico e previsão das CF e CE

Na figura seguinte, podemos observar como as CF se comportam em relação aos IO, de forma semelhante à análise feita na Figura 2.21, onde se analisa o valor do CAPEX, ou seja, o valor efetivamente despendido pela EDP Distribuição no que toca aos IO.

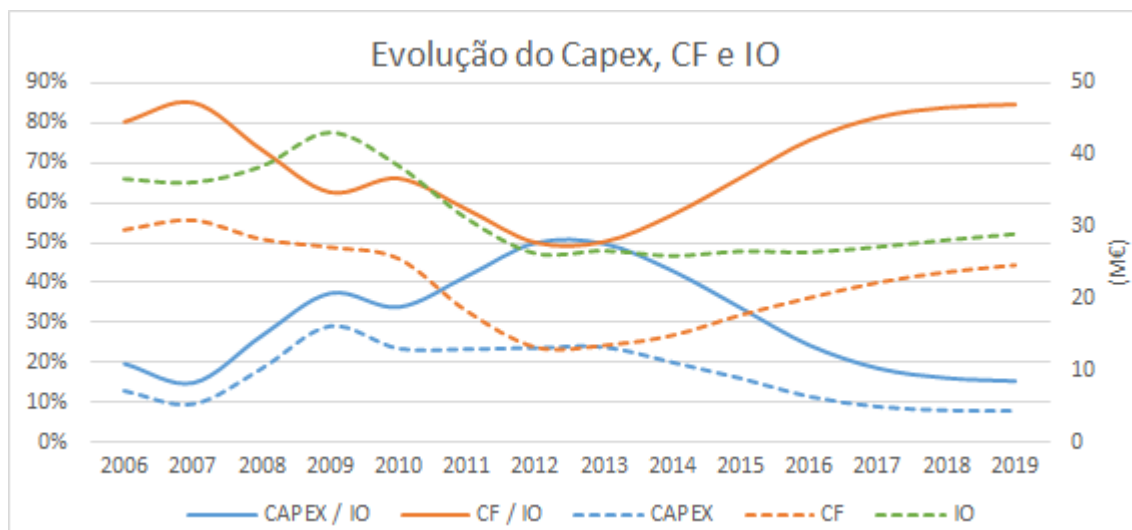


Figura 3.24 - Evolução do CAPEX das CF e dos IO decorrente da alteração na legislação

Como se observa no gráfico, as CF, em 2019, tal como acontece em 2007, estão muito próximas do valor dos IO. De notar ainda que o CAPEX/IO, desde a entrada da alteração na legislação em 2007, encontra-se em contínuo crescimento até 2013, data em que se dá uma nova alteração na legislação. Posteriormente a esta data, observa-se que o valor efetivamente despendido pela EDP Distribuição volta a representar uma proporção equivalente ao despendido anteriormente a 2007, como referido na subsecção 2.1.3..

Capítulo 4

Conclusões e Desenvolvimentos Futuros

Neste capítulo final são expressas as principais conclusões a extrair da pesquisa e do estudo desenvolvido, bem como os diversos obstáculos sentidos na sua execução. Para serem considerados em trabalhos futuros, na última secção, são enunciados um conjunto de sugestões possíveis para melhorar as estimações e os processos. Ainda nesta última secção são propostos vários complementos e modificações ao programa desenvolvido.

4.1 - Conclusões Finais

O principal objetivo de munir os ORD de uma ferramenta que lhes permita prever os IO necessários de incluir no PDIRD, para um horizonte temporal de cinco anos, nos quais se dá principal relevo às realizadas para o primeiro biénio, foram cumpridos com sucesso. É importante salientar que as previsões obtidas nas estimações dos IO encontram-se em conformidade com as previsões disponíveis para a economia portuguesa. Conseguiu-se diminuir de forma significativa a discrepância entre estimações dos modelos selecionados³. Não havendo outros exemplos semelhantes para comparação, percebe-se uma das maiores dificuldades do estudo desenvolvido, na medida em que se partiu praticamente do zero na definição de uma metodologia.

Quanto às diversas abordagens aplicadas aos dados de entrada, concluiu-se que a variação anual é compensatória apenas em variáveis que apresentem uma tendência constante, como se observa na estimação do IC, visto esta ser suscetível de fornecer estimações com picos elevados, saindo assim fora dos padrões definidos. Contrariamente, quando se utiliza a dessensibilização das entradas consegue-se limitar a influência dos picos existentes, aquando do treino de coeficientes, permitindo obter resultados mais consistentes, como se observou nas estimações das realizações e dos IO.

³ Em comparação com as estimativas realizadas pelo INESC Porto.

Conclui-se ainda que, mesmo usando variáveis com uma baixa relação com a saída, foi possível obter ME capazes de gerarem estimações com um erro relativamente baixo.

O escasso histórico disponível representa a grande limitação encontrada no desenvolvimento deste estudo, tanto a nível de processos de estimação a adotar, como de possíveis conclusões a retirar das suas estimações. Acrescente-se ainda o caso dos picos nos “anos de eleições autárquicas”, aparentemente inexplicáveis pelas restantes variáveis, mas como referido com um histórico tão reduzido estas “afirmações” podem não passar de especulações.

O desenvolvimento deste estudo mostrou-se bastante moroso, na medida em que foram necessárias explorar diversas alternativas e técnicas, de forma a encontrar a que melhor responde aos objetivos do trabalho. O programa desenvolvido mostrou ser uma mais-valia, na medida em que consegue diagnosticar e registar todos os ME possíveis de gerar em cerca de um minuto, o qual, segundo os procedimentos normais, levaria meses a conseguir-se.

4.2 - Desenvolvimentos Futuros

Visto que este é um estudo iniciado recentemente, existe ainda um longo caminho a percorrer, com inúmeras possibilidades e metodologias a serem testadas. No entanto, como existem prazos a cumprir, não é possível estendermo-nos a todas as possibilidades, pelo que não foi possível completar um dos objetivos do trabalho, isto é, a distribuição por UO dos IO.

Para a realização deste ponto, aconselha-se o recurso ao programa desenvolvido, trocando apenas as entradas que se encontram distribuídas por DRC, como o NNC, NNF e as realizações, pelo valor correspondente a cada DRC. Igualmente, troca-se a saída pelo histórico correspondente a cada DRC. Posteriormente, deve fazer-se uma distribuição proporcional dos valores estimados, para que a soma destes coincida com os valores globais dos IO.

O banco de variáveis disponíveis já constitui um número razoável e diversificado, pelo que não se acha compensatória a inclusão de muitas mais. Contudo, existem algumas alterações que se julgam compensatórias nos dados/variáveis utilizados, aquando do treino de coeficientes em processos de estimação, tais como:

- Reavaliar o histórico disponível para os IO, participações e realizações, onde se pretende moldar/justificar os picos nestas existentes;
- Compor um histórico das realizações com dados monetários, de forma a poder agrupar estas numa só variável;
- Constituir um histórico com o custo afeto a obras de grande escala, como por exemplo, a instalação de grandes indústrias ou urbanizações, semelhante ao realizado para a eletrificação da Serra da Serpa. Da mesma forma, incluir nesta variável as previsões para a realização deste tipo de obras, ao analisar quais as indústrias que se pensam implementar em Portugal ou projetos de grandes urbanizações, podendo estes dados serem distribuídos por DRC.

Quanto ao programa desenvolvido, deixam-se algumas sugestões capazes de melhorar a forma como os resultados são obtidos:

- No processo de estimação podem ser criadas hipóteses para o recurso a mais tipos de regressões. De forma semelhante, pode ser analisada qual a regressão que melhor relação entre as variáveis apresenta e conjugar na mesma estimação diversos tipos de regressão;
- Quanto à seleção dos melhores modelos, isto é, no processo de ordenação da MR, podem ser também conjugados com o atual critério, outros, como, por exemplo, a diferença entre o erro de treino e teste;
- Já no processo de eliminação de ME, aconselha-se conjugar com o atual critério um desvio máximo à estimação devolvida pela média dos melhores modelos, presentes na MR ordenada.

Referências

- [1] Decreto-Lei n.º 29/2006 de 15 de fevereiro. Diário da República n.º 33, Série I-A. Ministério da Economia e da Inovação. Lisboa.
- [2] Decreto-Lei n.º 215-A/2012 de 8 de outubro. Diário da República, n.º 194, Suplemento, Série I. Ministério da Economia e do Emprego. Lisboa.
- [3] Decreto-Lei n.º 215-B/2012 de 8 de outubro. Diário da República, n.º 194, Suplemento, Série I. Ministério da Economia e do Emprego. Lisboa.
- [4] Decreto-Lei n.º 172/2006 de 23 de Agosto. Diário da República, n.º 162, Série I. Ministério da Economia e da Inovação. Lisboa.
- [5] Saraiva, J. T., Fidalgo, N. e Costa, R. *Estimação de Investimentos Obrigatórios*. Unidade de Sistemas de Energia - INESC TEC. Guimarães, Novembro 2012.
- [6] INESC TEC. *Estimação de Investimentos Obrigatórios*. Proposta para a EDP Distribuição. Março, 2012.
- [7] Monteiro, C. *Redes Neurais*. Decisão, Optimização e Inteligência Computacional, Faculdade de Engenharia da Universidade do Porto, 2007
- [8] Cavaleiro, Diogo. *Portugal com segunda maior queda da UE no consumo de energia entre 2006 e 2012*.
[http://www.jornaldenegocios.pt/empresas/detalhe/portugal_com_segunda_maior_que_da_da_ue_no_consumo_de_energia_entre_2006_e_2012.html, acedido em 3 de março de 2014]
- [9] Instituto Nacional de Estatística. *Dados Estatísticos*.
[http://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_base_dados&contexto=bd&selTab=tab2, acedido entre Março e Junho de 2014]
- [10] Instituto Nacional de Estatística. *Índice de Preços no Consumidor Base 2002 - Nota Metodológica*. Fevereiro, 2013
- [11] Gabinete de Estratégia e Estudos (GEE). *Previsões Macroeconómicas/ Macroeconomic Forecasts*. GEE, Ministério da Economia. Fevereiro 2014. [www.gee.min-economia.pt/en/?cfl=29099]
- [12] Federação Portuguesa da Indústria da Construção e das Obras Públicas. *Construção termina 2013 com nova queda de 15%*. Fevereiro, 2014.

- [<http://www.fepicop.pt/index.php?id=50&tbl=noticias>, acedido em 15 de março de 2014]
- [13] Federação Portuguesa da Indústria da Construção e das Obras Públicas. *Construção inicia 2014 com tímidos sinais positivos*. Maio, 2014.
[<http://www.fepicop.pt/index.php?id=51&tbl=noticias>, acedido em 25 de maio de 2014]
- [14] ERSE Entidade Reguladora dos Serviços Energéticos. Despacho nº 18 993-A/2005, a 31 de Agosto. Diário da República, 2.ª Série - Nº 167. ERSE, 2005.
- [15] ERSE Entidade Reguladora dos Serviços Energéticos. Despacho nº 12 741/2007, a 21 de Junho de 2007. Diário da República, 2.a série — Nº 118. ERSE, 2007.
- [16] ERSE Entidade Reguladora dos Serviços Energéticos. Despacho nº 6402/2011, a 14 de Abril de 2011. Diário da República, 2.ª série — N.º 74. ERSE, 2011.
- [17] ERSE Entidade Reguladora dos Serviços Energéticos. *Ligações às redes de energia elétrica*. Setembro 2011.
- [18] ERSE Entidade Reguladora dos Serviços Energéticos. *Regulamento de Relações Comerciais do Setor Elétrico*. Outubro 2012.
[http://www.erse.pt/pt/electricidade/regulamentos/relacoescomerciais/Documents/RR_C_articulado_2012.pdf, acedido em 12 de março de 2014]
- [19] ERSE Entidade Reguladora dos Serviços Energéticos. Diretiva n.º 18/2012, a 8 de novembro de 2012. Diário da República, 2.ª série — N.º 216. ERSE, 2012
- [20] ERSE Entidade Reguladora dos Serviços Energéticos. *Ligações às redes de energia elétrica de instalações consumidoras (em vigor a partir de Maio de 2013)*. Novembro 2012.
- [21] Fidalgo, N. e Monteiro, C. *Aula 2 - Modelos de regressão*. Técnicas para Previsão, Faculdade de Engenharia da Universidade do Porto, 2013.
- [22] Suporte Office. *Pearson*. [<http://office.microsoft.com/pt-pt/excel-help/pearson-HP005209210.aspx>, acedido em 27 de Fevereiro de 2014]
- [23] Suporte Office. *Obter informações sobre como prever e mostrar tendências em gráficos*. [<http://office.microsoft.com/pt-pt/excel-help/adicionar-alterar-ou-remover-uma-linha-de-tendencia-num-grafico-HP010342158.aspx#BMrSquaredvalue>, acedido em 27 de fevereiro de 2014]
- [24] Azevedo, F. e Machado, J. T. *Analysis of Electricity Markets Using Multidimensional Scaling*. 9th International Conference on the European Energy Market, Florença-Itália, 10-12 de maio, 2012
- [25] Saraiva, J. T., Azevedo, F., Fidalgo, N., Fonseca, N., Gonçalves, R., Costa R. N. *Estimação de investimentos obrigatórios a realizar pela EDP - Distribuição*. INESC Porto, Dezembro 2012.